Ethics and Information Technology
https://doi.org/10.1007/s10676-021-09614-x

ORIGINAL PAPER

Ethics of automated vehicles: breaking traffic rules for road safety

Nick Reed[1] · Tania Leiman[2] · Paula Palade[3,4] · Marieke Martens[5] · Leon Kester[6]

Accepted: 2 September 2021

Abstract
In this paper, we explore and describe what is needed to allow connected and automated vehicles (CAVs) to break traffic rules in order to minimise road safety risk and to operate with appropriate transparency (according to recommendation 4 in Bonnefon et al., European Commission, 2020). Reviewing current traffic rules with particular reference to two driving situations (speeding and mounting the pavement), we illustrate why current traffic rules are not suitable for CAVs and why making new traffic rules specifically for CAVs would be inappropriate. In defining an alternative approach to achieving safe CAV driving behaviours, we describe the use of ethical goal functions as part of hybrid AI systems, suggesting that functions should be defined by governmental bodies with input from citizens and stakeholders. Ethical goal functions for CAVs would enable developers to optimise driving behaviours for safety under conditions of uncertainty whilst allowing for differentiation of products according to brand values. Such functions can differ between regions according to preferences for safety behaviours within that region and can be updated over time, responding to continual socio-technological feedback loops. We conclude that defining ethical goal functions is an urgent and necessary step from governmental bodies to enable the safe and transparent operation of CAVs and accelerate the reduction in road casualties they promise to achieve.

Keywords Automation · Ethics · Safety · Driving · Vehicles

Introduction

Vehicle automation is anticipated to improve road safety based on the premise that this will eliminate common mistakes and misjudgements made by human drivers (e.g. Morando et al. 2017; Fagnant & Kockelman, 2015; Kyriakidis et al., 2019). In self-published research (Scanlon et al., 2021), automated vehicle developer, Waymo, claimed that their automated driving systems can avoid most collisions and reduce the severity of unavoidable crashes within specific operational design domains[1] (ODDs). In simulated reconstructions of 72 fatal incidents with human drivers, Waymo virtually replaced the human driver whose manoeuvre resulted in the fatal collision with Waymo's automated driving systems. Their simulations showed that this substitution would have prevented all fatal collisions.

However, despite frequent claims from industry regarding potential possible advantages, automated driving in its various forms can still lead to collisions (e.g. Dutch Safety Board, 2019; NHTSA, 2017; NTSB, 2018, 2019). As technology evolves, new risks arise due to mode confusion, loss of situational awareness of the person inside the vehicle, transition of control problems and automation surprise (e.g. Carsten & Martens, 2019; Cummings & Ryan, 2014; Martens & Van den Beukel, 2013). Irrespective of human-system interaction issues, the complexity of traffic, the interaction with other road users and vulnerable road users, unexpected

✉ Nick Reed
nick@reed-mobility.co.uk

1 Reed Mobility, Wokingham, UK
2 Flinders University, Adelaide, Australia
3 Jaguar Land Rover, Abbey Road, Whitley, Coventry CV3 4LF, UK
4 University of Bradford, Bradford, UK
5 TNO, Helmond, The Netherlands
6 TNO, The Hague, The Netherlands

[1] Operational design domain is the operating conditions under which a given driving automation system or feature thereof is specifically designed to function (SAE, 2018).

Published online: 11 October 2021

events, system limitations or failures may all produce new types of collision risk for automated vehicles.

A recent report for the European Commission (Bonnefon et al., 2020) included twenty recommendations addressing the ethics of connected and automated vehicles (CAVs[2]), covering topics such as road safety, privacy, fairness, explainability and responsibility. With respect to road safety, the report suggests 'a minimal requirement for manufacturers and deployers is to ensure that CAVs decrease, or at least do not increase the amount of physical harm incurred by users of CAVs or other road users that are in interaction with CAVs, compared to the harm that is inflicted on these groups by an appropriately calculated benchmark based on conventional driving' (Recommendation 1, Bonnefon et al., 2020) and suggested that the introduction of CAVs requires careful consideration of the circumstances in which they might be permitted not to comply with all applicable traffic rules (Recommendation 4, Bonnefon et al., 2020).

Although possibly counterintuitive, in some situations, departure from strict compliance may be required to minimise the risk or severity of harm. This is reflected in existing legal provisions by the use of terminology which envisages the use of discretion by human drivers, such as driving with 'reasonable consideration' and 'due care'. Human drivers exercise this discretion based on experience, training and a general understanding of the road environment.

While it is difficult for CAVs to be programmed to exercise similar discretionary behaviour, requiring CAVs to comply strictly with traffic rules may not necessarily achieve optimal road safety. However, allowing exceptions to a traffic rule would require the CAV to evaluate the safety risks associated with compliance and non-compliance and where a non-compliant behaviour is deemed to pose a lower overall risk of harm, then to select that non-compliant behaviour. This task is far from trivial.

For human drivers, behaviours are shaped by laws, rules and guidelines to promote road safety—typically presented as a national reference guide [e.g. UK: The Highway Code (Driver and Vehicle Standards Agency, 2015); Netherlands: Road Traffic Act (WVW, 1994) and Traffic Regulations and Road signs (Rvv, 1990); France: Code de la route (Legifrance, 2021)]. Knowledge of traffic law and regulations are a critical element in the licence acquisition process for human drivers.

CAVs could be programmed to follow digital versions of statutory road rules and the Highway Code (or equivalent for its jurisdiction of operation). However, as Prakken (2017) has discussed, there are challenges in making CAVs conform to traffic law. The complexity of domains in which CAVs are expected to operate makes it impossible to anticipate or formulate explicit rules for every potential eventuality—environmental conditions, traffic and encounters with other road users will vary dramatically between domains and over time in any one domain.

However, any approach that assumes a CAV could 'deduce' normatively correct behaviour through exposure to a large number of training cases would need to overcome three extremely challenging practical difficulties:

(1) Collecting a sufficient quantity and quality of scenarios to allow the right behaviours to be derived, especially since traffic collisions tend to happen in the tail of the distribution of driving and are therefore rare. No training data set can exhaust all possibilities.
(2) In the unlikely event that this could be achieved, a CAV will not derive the values or ethical principles as to *why* any specified decisions or behaviour should be adopted, and therefore cannot develop ethical principles to apply when confronted by new situations.
(3) An automated system that has 'deduced' driving behaviour from training examples cannot 'explain' or 'justify' its decisions or actions. This 'opacity, connectivity and autonomy' (European Parliament, 2020) may be problematic if a manufacturer is required to explain specific behaviour in case of an incident or where civil or criminal liability is disputed (see also recommendation 4 of Bonnefon et al, 2020). In fault based tort law systems (European Parliament, 2020) injured persons claiming compensation for road trauma might be required to prove negligence, or establish precise causative links between that negligence and their injuries or damage. Persons charged with a criminal breach of traffic rules might dispute that they committed the alleged act, or that they did so with the necessary intention, or both.

In this paper, we describe why current traffic rules are unsuitable for CAVs and why making new CAV traffic rules will be insufficient to maximise road safety. We also explain how CAVs might break traffic rules to minimise road safety risk and provide transparency in the event of a collision (according to recommendation 4 in Bonnefon et al, 2020) and why we believe that ethical goal functions are needed to deliver safe driving behaviour of CAVs.

Although 'CAV' can cover a range of automation levels, in this article we use it broadly to refer to vehicles with Conditional (SAE Level 3), High (SAE Level 4) and Full (SAE Level 5) driving automation. While we acknowledge significant discussion exists regarding the SAE levels and

[2] Bonnefon et al. (2020) defined CAVs as vehicles that are both connected and automated and display one of the five levels of automation according to SAE International's J3016 standard, combined with the capacity to receive and/or send wireless information to improve the vehicle's automated capabilities and enhance its contextual awareness.

how specific levels will function (latest SAE J3016 update, April 2021), in this article we do not use 'CAV' to refer to vehicles characterised as SAE Level 1 or 2 as the human driver is still monitoring the road and responsible for the dynamic driving task (DDT). We refer instead to vehicles that have Automated Driving Systems (ADS), described by SAE as: 'The hardware and software that are collectively capable of performing the entire DDT on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD). Since most future vehicles will also be connected, we specifically use the term CAV.

We begin by reviewing current traffic laws and considering how those might apply to CAVs. Acknowledging Bonnefon et al.'s (2020) Recommendation 4, we examine two specific traffic situations (exceeding the speed limit; mounting the pavement) in which a CAV might be allowed to 'breach' traffic rules to minimise road safety risk. We briefly canvas opinions from law enforcement organisations, legal experts, transportation researchers (human factors and traffic safety experts), artificial intelligence[3] (AI) experts, industry and non-experts (via the consultation of the Law Commission) and draw on publicly available consultation responses from organisations responding to specific questions on traffic rule compliance.

We close by analysing how CAVs might determine when it could be deemed ethically acceptable to deviate from the traffic rules in the interests of road safety, finding that, even based on the two traffic situations evaluated within the paper, traditional AI approaches to determining CAV behaviour[4] are inadequate for this task. We conclude that while specific rules and regulations (symbolic reasoning) for CAVs are important, ethical goal functions are needed to resolve situations where traffic rules are insufficient.

Traffic rules and regulations

Traffic rules and regulations help to maintain order on our roads, thus increasing safety for all road users. While these differ somewhat between jurisdictions, UN conventions on road traffic provide some international coherence [Geneva (UN, 1949); Vienna (UN, 1968, 2016)]. A careful and competent human driver is presumed, as a minimum, to comply with the road rules, and to take reasonable care when driving not to cause harm to others—but there is little guidance beyond that. As illustrated below, mapping similar notions of competence to CAVs poses different challenges.

Breaches of road traffic rules may be summary offences, dealt with in the lowest courts (Road Traffic Offenders Act, 1988). Breaches of some road rules (such as exceeding specified speed limits) (Road Traffic Regulation Act, 1984) may be offences of strict liability—where only externally observable behaviour has to be proved (e.g. the car was travelling at a particular speed in a particular speed zone), and any intention or recklessness by drivers is not relevant. Breaches of other rules may require proof of more nuanced factors, particularly offences that require proof of driving without due care or attention or without reasonable consideration for other persons using the road, (Road Traffic Act, 1988s.3) failure to act reasonably, failure to be in proper control of the vehicle (all of which envisage a human driver making decisions and may require proof of intention or recklessness), or where defences are available (e.g. in case of emergency etc.).

Careless driving may include a momentary lapse of concentration or misjudgement, loss of control due to speed or insufficient attention to road conditions. When this happens at low speeds, in situations where other people or property are not impacted, or when it is not observable by others, charges of breaching road rules are very unlikely to be laid—especially if the behaviour has not been observed directly by the police, or is not reported to them, or result in an informal warning to the human driver. Human drivers regularly have these momentary lapses, usually with no legal consequences. However, where driving falls much further below the standard expected of a competent careful driver, and this is externally observable or impacts on others, the likelihood of a driver being charged by traffic authorities with an offence increases.

In the UK, the Road Traffic Act (1988) provides for offences in relation to dangerous driving (see Sects. 1, 1A, 2 and 2A), careless and inconsiderate driving (see Sects. 2B, 3, 3ZA, 3A).

Section 2 provides:

A person who drives a mechanically propelled vehicle dangerously on a road or other public place is guilty of an offence.

Section 2A defines 'dangerous driving':

(1) For the purposes of sections 1 and 2 above a person is to be regarded as driving dangerously if (and, subject to subsection (2) below, only if)—

  (a) The way he drives falls far below what would be expected of a competent and careful driver, and
  (b) It would be obvious to a competent and careful driver that driving in that way would be dangerous.

[3] In accordance with Russell & Norvig (2009), we define AI as rational agents that use data and computation to determine actions that achieve the best expected outcome.

[4] By 'traditional', we mean approaches based on deep learning to determine optimal outcomes.

(2) A person is also to be regarded as driving dangerously for the purposes of sections 1 and 2 above if it would be obvious to a competent and careful driver that driving the vehicle in its current state would be dangerous.
(3) In subsections (1) and (2) above "dangerous" refers to danger either of injury to any person or of serious damage to property; and in determining for the purposes of those subsections what would be expected of, or obvious to, a competent and careful driver in a particular case, regard shall be had not only to the circumstances of which he could be expected to be aware but also to any circumstances shown to have been within the knowledge of the accused.
(4) In determining for the purposes of subsection (2) above the state of a vehicle, regard may be had to anything attached to or carried on or in it and to the manner in which it is attached or carried.

Section 3ZA defines 'careless and inconsiderate driving':

(1) …
(2) A person is to be regarded as driving without due care and attention if (and only if) the way he drives falls below what would be expected of a competent and careful driver.
(3) In determining for the purposes of subsection (2) above what would be expected of a careful and competent driver in a particular case, regard shall be had not only to the circumstances of which he could be expected to be aware but also to any circumstances shown to have been within the knowledge of the accused.
(4) A person is to be regarded as driving without reasonable consideration for other persons only if those persons are inconvenienced by his driving.

Failing to stop at a red light, failing to give way, significantly exceeding speed limits, driving on the wrong side of the road, failing to stop at the scene of an accident or when asked to do so by the police, or driving under the influence of alcohol or drugs are other behaviours that might (depending on circumstances) give rise to a charge of dangerous driving, in addition to specific lesser charges.

However, while traffic and vehicle regulatory offences may comprise as much as one third of the matters before criminal courts (Australian Bureau of Statistics, 2021), the experience of most human drivers suggests they are not charged every time they infringe traffic rules slightly, especially when not observed by police or no collision is caused. Traffic authorities exercise discretion over whether to prosecute the human driver when breaches are observed. Minor breaches may not be prosecuted, with human drivers instead being warned or counselled by police. Even where gross breaches of the standard expected of competent and careful human drivers occur, if this is not observed by others or by authorities, prosecution is unlikely.

CAVs are fundamentally different with systems recording critical aspects of the vehicle's operation including location, speed, acceleration and braking. All instances of exceeding applicable speed limits (even by small amounts) are thus identifiable, potentially in real time, even if not observed externally by traffic authorities or other road users. This poses challenges as to whether charges should be laid for every infringement, the impact this might have on the workload of judicial bodies and what defences, if any, might apply to avoid criminal liability.

Even a rule as apparently simple as a speed limit presents challenges in this regard. In the UK, it is an offence to drive 'a motor vehicle on a road at a speed exceeding a limit imposed' by statute (Road Traffic Regulation Act, 1984, Sect. 89). The UK's Highway Code notes:

> You must not drive faster than the speed limit for the type of road and your type of vehicle. The speed limit is the absolute maximum—it doesn't mean it's safe to drive at this speed in all conditions.

In addition, specific speed limits might be imposed on vehicles of a particular class (e.g. Road Traffic Regulation Act, 1984, Sect. 86), and emergency services vehicles are exempted from speed limits when used for emergency purposes (e.g. Road Traffic Regulation Act, 1984, Sect. 87).

Human drivers are thus always required to make decisions about what driving speed is safe in all circumstances (up to the speed limit). Choosing to drive at the maximum permissible speed will not result in the offence of exceeding the speed limit but could still amount to driving without reasonable consideration for other persons using the road. Human drivers can exercise their discretion to proceed through an intersection against a red light to make way for an ambulance, fire or police vehicle or where other emergency situations arise, even though the road rules require vehicles to stop when traffic lights are red.

Again, such discretion is problematic for CAVs. It may be relatively simple to program a CAV to drive in the correct lane, at the designated speed limit, and maintain an appropriate distance from other vehicles. However, even this compliant behaviour may not avoid a charge of driving without due care and attention given other external circumstances such as poor weather conditions that make driving at the designated speed limit unsafe. CAVs may be programmed to give way to emergency vehicles but this may not resolve the dilemma of how to code for competing priorities or when CAVs should operate in ways not ordinarily prescribed by the rules. What value should CAV algorithms assign to the various competing safety priorities that may exist given on any road that is not a closed system?

For safe operation, a CAV must behave in ways that are predictable for other road users. When vehicle behaviour cannot be reliably predicted, this may expose manufacturers, fleet owners and operators (and possibly even private vehicle owners/operators) to liability for traffic offences, and also potential for civil liability if harm or damage is caused to other road users as a result. If other road users expect that CAVs will always stop for a red light, or travel at a particular speed, and a CAV behaves differently, then arguably this might amount to acting without reasonable consideration for other persons using the road who should be entitled to be able to predict driving behaviour. This is made even more complex if unpredictability in CAV behaviour is not always apparent to other human road users. The importance of predictability is confirmed by collision statistics. Goodall (2021) noted automated vehicles were 4.8 times more likely to be struck from behind than human-driven vehicles in a naturalistic driving study. Goodall suggests that automated vehicles' behaviour concerning where and when to stop or remain stopped at intersections was a likely contributor to this elevated risk.

Evidence that a vehicle has breached road rules may be important evidence in any civil claim brought to recover compensation for any injury or harm caused to the vehicle's occupants, other road users or property. However, given different standards of proof in the criminal and civil context, it may be enough to impose civil liability if a failure to take reasonable care is proved on the balance of probabilities (i.e. more likely than not) rather than on the criminal standard of beyond reasonable doubt. Vehicle data is increasingly valuable here (even for human-driven vehicles), and CAV data will be critical in ascertaining exactly how a vehicle has behaved on any particular occasion. Where injuries or losses are significant (and so stakes are high), it should be anticipated that prospective claimants in any legal proceedings will also seek to interrogate operational settings including any algorithms determining the vehicle takes one course of action over another. As recently noted, 'documentation and traceability, [and] transparency' are critical for such high-risk AI (European Commission, 2021).

Potential liability for breaches of road rules may thus be very difficult to assess prospectively, particularly in the first cases to come before traffic authorities, the police, or the courts. Such uncertainty may inhibit the broader adoption of CAV technology, in turn delaying potential safety gains.

Industry, expert and public opinion

To explore the issue of CAV compliance with road rules, we focused on two situations:

1. Exceeding the speed limit
2. Mounting the kerb

Selection of these situations builds on earlier work of the Law Commission of England and Wales that was tasked with reviewing the laws for operating CAVs on public roads. It has published three open consultation papers, the first of which invited respondents to answer questions about the behaviour of CAVs with respect to these specific two offences (Law Commission, 2018). Where permitted by the 178 respondents to the consultation,[5] responses to the Law Commission's questions were published (Law Commission, 2019). We explored how responding organisations anticipated CAVs should behave in these two situations.

We also informally asked various industry experts in the authors' network in the fields of CAVs, human factors and road safety to share their views on how to respond to the challenges of programming CAVs to comply with such traffic rules. Our intention was to understand whether there was agreement across people working in this field about how this issue could be resolved. If it were possible to make rules that could adequately govern CAV behaviour, then the additional ethical goal functions for which we advocate in our introduction may be unnecessary.

Speeding

Speed is a critical risk factor increasing the risk and severity of crashes (Elvik et al., 2019). Speed limits on public roads provide drivers (and CAVs) with enforceable guidance about maximum permissible speed, yet speeding is commonplace (e.g. Department for Transport, 2020; Quimby et al., 2005). Against this background, respondents to the Law Commission consultation paper (Law Commission, 2019) varied in their responses as to how CAVs should behave with respect to speed limits.

Arguments against allowing CAVs to exceed the speed limit

Some respondents claimed there were no circumstances in which speeding by a CAV should be allowed, reflecting how prescribed speed limits account for the laws of physics [kinetic energy of an impact, perception (detection) and reaction times] and aim to increase crash survivability. Also, concerns were raised that any tolerance would become the de facto speed limit.

Some respondents claimed that speed limit tolerances were a legacy of the inaccuracies of older speedometers, limitations in enforcement technology, short periods of human inattention or adaptation to lower speeds. Arguably,

[5] Respondents included vehicle manufacturers, CAV developers, industry bodies, national and sub-national transport organisations, charities, technology companies, police forces, law firms and private individuals (Law Commission, 2019).

such tolerances are not applicable for CAVs since speeds can be monitored continuously and accurately.

Also, respondents claimed any argument for excess speed in terms of safety is flawed logic. Any action that requires excessive speed should either not have been started in the first place or should be abandoned. These respondents suggest compliance with speed limits is expected, with some even arguing for a further reduction of limits in built-up areas.

Arguments for allowing CAVs to exceed the speed limit

Some respondents claimed they can imagine conditions in which it may be safe to allow CAVs to exceed the speed limit. However, none thought CAVs should be permitted to travel markedly faster than the speed limit (at least not for now). Arguments made included setting acceptable tolerances and limiting this to specific cases and appropriate conditions. This begs the questions as to what are those cases or conditions and whether these can be determined in advance or in real time as the vehicle encounters the situation.

Some respondents considered that avoidance of collisions by overtaking quickly was a circumstance where speeding may be necessary for safety. However, respondents were not prepared to support speeding simply at the request of the vehicle user, operator or occupant.

Speed differential is a significant factor in collision severity. Some respondents argued that CAVs in mixed traffic settings should be allowed to speed above the threshold to adopt similar speeds as other vehicles to reduce frustration amongst other road users (potentially leading to unsafe overtaking manoeuvres by other road users), with the suggestion that this would improve road safety.

This leads to the question of how to define conditions in which speeding might be permitted. Some respondents proposed general principles; others formulated very specific descriptions of what needed to be done in a specific scenario. Proposed general principles for allowing CAVs to exceed the speed limit included:

- Since applicable speed limits are location specific, already determined by risks of that environment (e.g. road contours, pedestrians, crossing traffic, adjacent activities such as schools, etc.), any determination of the amount by which CAVs would be allowed to exceed any applicable speed limit cannot be a fixed number, but must vary depending on those circumstances.
- The period of time for which exceeding the speed limit is permissible should be limited, with a maximum of 5 or 10 s.

Proposed specific situations where speeding would be safer than strictly adhering to the speed limit included:

- Finishing an overtaking manoeuvre at higher speed in order to avoid colliding with oncoming traffic;
- Transitioning smoothly from a higher limit to a lower limit (instead of braking sharply potentially causing unpredictable behaviour).

Mounting the kerb

The UK's Road Traffic Act, 1988 Sect. 34 provides.

> (1) Subject to the provisions of this section, if without lawful authority a person drives a mechanically propelled vehicle—
>
> (a) on to or upon any common land, moorland or land of any other description, not being land forming part of a road, or
> (b) on any road being a footpath, bridleway or restricted byway, he is guilty of an offence.
>
> (2) ...
> (2A) ...
> (3) It is not an offence under this section to drive a mechanically propelled vehicle on any land within fifteen yards of a road, being a road on which a motor vehicle may lawfully be driven, for the purpose only of parking the vehicle on that land.
> (4) A person shall not be convicted of an offence under this section with respect to a vehicle if he proves to the satisfaction of the court that it was driven in contravention of this section for the purpose of saving life or extinguishing fire or meeting any other like emergency.
> (5) …
> (6) ...
> (7) …
> (8) ...

The Law Commission (2018) noted that there is very little case law on what the exception noted in subsection (4) covers. However, they recognised three situations in which an otherwise careful and law-abiding driver might mount the kerb in an urban environment. Firstly, to avoid a collision (especially to avoid injuring another human road user); secondly, to permit an emergency vehicle to pass (even though The Highway Code explicitly notes that drivers should avoid mounting the kerb in such a situation) and thirdly, to allow another vehicle to pass in a narrow street. The Law Commission consultation paper (Law Commission, 2019) asked respondents their views on whether CAVs should be allowed to mount the pavement in each of the three situations.

Arguments against allowing CAVs to mount the kerb

A minority of respondents suggested that CAVs should never mount the pavement, under any circumstance, citing concerns over risks to vulnerable road users (children, elderly, disabilities, visually and hearing impaired) and their ability to avoid an approaching CAV.

However, many argued that mounting the pavement at speed should never be permitted, identifying risks of losing control of the vehicle as a result of the impact with the kerb or a tyre bursting, and risks that street furniture such as lamp posts could injure the vehicle occupants.

An alternative viewpoint suggested that a human 'user-in-charge' (Law Commission, 2018) should take over control of a CAV to perform any manoeuvres necessary to allow an emergency vehicle to pass or to enable traffic flow that would require departure from the road. It was felt that they could assess the benefits and risks of any decision and would be responsible for any deviation from the relevant law. However, this introduces new challenges in promptly and accurately detecting the situations where this switch to human control becomes necessary and safely achieving the transition (a much-studied issue: e.g. Lu et al., 2016; Zhang et al., 2019 for a meta-review).

Arguments for allowing CAVs to mount the kerb

A majority of the respondents considered that avoiding a collision could be sufficient to warrant a CAV behaving in this way but only as a last resort and that consideration must be given to ensuring safety for pedestrians and cyclists. The Society of Motor Manufacturers and Traders UK (SMMT) proposed that the dilemma of whether to mount pavements could be dealt with through the redesign of the road environment rather than extending the automated driving system's ODD to include pavements. They further proposed that until Emergency Vehicle Warning (vehicle-to-vehicle (V2V)) communication becomes standard, mounting the pavement should be regarded as an acceptable temporary solution permitted as a last resort.

A small majority thought mounting the kerb would be acceptable in order to allow emergency vehicles to pass (at low speeds). About a third of the respondents thought that CAVs should be allowed to mount the kerb to enable traffic flow and prevent gridlocks. This implies that optimisation of CAV behaviour goes beyond safety to encompass other utilities (traffic flow) while maintaining road safety levels. Others argued safety cannot be traded for convenience.

An interesting perspective was that CAV developers should be able to program CAVs such that, if the pavement is within the ODD of the vehicle, the car should be able to resolve conflict situations, using the pavement only when necessary. It is not clear how this would happen in the absence of adequate training data, and given the infinite unpredictability of situations, the types of surfaces and the vehicle capabilities that might be encountered.

Some respondents suggested that CAVs could apply a hierarchy that, firstly, evaluates the probability and severity of collision by remaining on the regular road against the likelihood of negative outcomes if mounting the pavement. If the vehicle is at relatively low speed and no hazards are detected on the pavement, the CAV may deviate away from the regular road to avoid a collision.

For other utilities, for instance aiding traffic flow efficiency by allowing another vehicle to pass on a narrow street, mounting the kerb can be managed at very low speed (walking pace) provided the pavement surface is improved (e.g. paved or gravel) and the CAV can determine that the region is clear of hazards. This means not only avoiding other road users but also not acting in a manner that could be perceived as threatening (e.g. by manoeuvring closely to a pedestrian or person sitting outside a street cafe).

These arguments collide again with the challenges identified by Prakken, discussed above. Whilst the law offers leeway in allowing a driver to mount the pavement in an emergency, the critical point is that the CAV must determine whether situations it encounters constitute an emergency and whether it would therefore be correct and justifiable to mount the pavement. Again, with all of the inherent variation in driving conditions, it would be impossible to pre-determine every set of circumstances for which mounting the pavement was an appropriate action and no training data set can exhaust all possibilities.

Furthermore, CAVs cannot 'deduce' a normatively correct 'understanding' as to why any specified decisions or behaviour should be adopted from a human perspective. This also means that CAVs cannot develop underlying ethical principles to apply in new situations, even if sufficiently large training data were available. They cannot 'explain' or 'justify' decisions or actions from an ethical point of view, which is problematic where civil or criminal liability is disputed. We explore this further below by considering ethical goal functions.

Proposals for CAVs to be able to mount the kerb in the case of making room for an emergency vehicle or avoid hitting an obstacle on the road assume that:

- Mounting the kerb presented less risk to safety than continuing on the roadway.
- A human driver would be justified in mounting the kerb in their vehicle in similar situations vehicle.
- A CAV's capabilities (for instance, detecting specific road surface, kerbs, infrastructure, pedestrians) could support safe navigation from the road onto the footpath/sidewalk and other safe operation in all circumstances.

- The CAV mounted the kerb at low speed, resembling speeds used when parking.
- Manoeuvres to rejoin the road would be managed manually (by a remote or onboard operator).

These assumptions may not be valid.

How can CAVs determine ethical rule breaking behaviours?

Our analysis revealed a lack of unanimity in preferences over CAV rule compliance, even for the relatively simple scenarios of exceeding the speed limit and mounting the kerb. It is therefore difficult to grasp how CAVs, guided by AI systems, could resolve such uncertainties. One type of machine learning used to develop CAVs is deep learning; a data driven approach in which CAVs 'learn' to recognise and respond to objects and environmental cues from exposure to many (real or simulated) situations on the road. In addition, symbolic algorithms are used to keep the behaviour within explicitly expressed boundaries such as those formulated by the law. This approach is known as neuro-symbolic hybrid AI (Marcus, 2020).

There are several challenges to this hybrid approach. For example, deep learning works very well for situations in which there are a lot of examples from which to develop associations, like object recognition from an a priori known set of possible objects. This is also true for low-level vehicle control (i.e. normal traffic behaviour) where relatively simple tasks such as lane-keeping and speed compliance can be learned. By contrast, it typically does not work well when encountering unknown objects and unusual and/or complex situations not covered by previous training datasets. For this reason, a human-defined symbolic layer is required to supervise the sub-symbolic deep learning algorithm and to make its outputs intelligible to humans. However, the symbolic approach needs a well-defined mathematical specification of what is 'good' or 'bad' behaviour; it needs ethics. Laws, rules and regulations provide an initial basis for this determination but in many complex situations, they are insufficient, particularly for moral dilemma cases where there is no absolute or clear 'good' or 'bad' behaviour.

So, how could a CAV, from a systems engineering perspective, determine when society would expect it to depart from absolute compliance with explicit rules in the interests of safety and how it should respond to emergencies or take other action regarded as appropriate in all reasonably foreseeable circumstances? Inevitably, this raises ethical questions about the relative value that society places on the responses possible in any given situation.

Analysis of state-of-the-art AI shows that, due to the lack of understanding of explanatory knowledge relevant in human morality, current systems can neither independently 'learn' to derive the ambiguous human values from human behaviour or human feedback nor apply them to new situations (Aliman, 2020). Thus, while in theory, a CAV could apply risk assessment and risk management algorithms that calculate optimal behaviour taking situational uncertainties and finite computational capabilities of the vehicle into account, for it to succeed in practice, one needs human-crafted mathematical heuristics encoding what behaviours are considered 'good' or 'bad'. In order to implement the required type of heuristic modelling, a so-called ethical goal function is needed, which describes the goal of any advanced AI system, and is an approximation of human values (Aliman & Kester, 2019). As discussed above such a goal function, also called utility function, cannot be 'learned' by the automated system. It must be constructed and defined by humans. This will involve significant complexity but will also give society control to define road safety as the primary utility function.

Note that, while for AI in ethical high-stake contexts, it has been shown that the classical utilitarianist and consequentialist utility functions face safety-relevant impossibility theorems (Eckersley, 2018), the ethical goal functions to which we refer to are not subject to these impossibility theorems. In fact, they are instead crafted within a novel non-normative meta-ethical framework denoted augmented utilitarianism (AU) and are formulated at a higher abstraction level which allows for a better heuristic modelling of morality (Aliman, 2020). While utilitarianism merely focuses on consequences of actions affecting 'the patient',[6] which means it considers a snapshot of the outcome of actions, AU-based ethical goal functions extend beyond that and facilitate the joint specification of multifaceted parameters for agent, action, perceiver and 'the patient' (which is in this case all road users that could be affected). In short, under AU, utility functions are not restricted to a snapshot of the final outcomes. Instead, they cover a time integral of a moral event (evolving over time) seen through the lens of a diversity of parameters which allows for a flexible heuristic moral meta-model for meaningful AI control in pluralistic societies (Aliman et al., 2019). Moreover, these AU-based functions are meant to be updatable-by-design. As opposed to classical utilitarianism, AU is meta-ethical and non-normative. It acts as a descriptive, explanatory and assistive framework providing a generic human-centred scaffold left blank. In this way, a representation of society can dynamically harness updatable AU-based ethical goal functions to specify which parameters matter to people and how much.

Unfortunately, ethicists, lawyers and legislators have to date not provided such ethical goal functions. However, we argue that without such functions (and in the absence of a

[6] Term used in Moral Theory.

complete overhaul of traffic rules to eliminate elements of nuance and discretion), adoption of CAVs for use in public and mixed operational domains will remain an unrealistic objective. Generally, in ethical high-stakes cases where the designer is not the central moral authority, a meta-ethical programming approach (of which AU represents an example) is required (Wernaart, 2021). If no ethical goal function is developed, decisions as to how CAVs should behave in challenging situations would be left to CAV manufacturers and deployers. CAV behaviours in challenging traffic situations could vary across individual brands or vehicle types and may not align with the values of society more widely. Discussions with CAV developers indicate they want clear guidance on how to do this rather than being left to make these decisions themselves. The absence of democratically developed ethical goal functions and uncertainty regarding whether behaviour of particular CAVs would be considered 'correct' may result in stagnation of CAV deployment.

Another solution discussed in the literature is to leave the responsibility for ethical decision-making with the user of a CAV. Contissa et al. (2017) proposed an 'Ethical Knob'—a device by which a vehicle occupant could customise the ethical principles adopted by the CAV according to their own personal preference. They suggest three modes: altruistic (preference to protect third parties), impartial (equal importance given to passenger(s) and third parties) and egoistic (preference to protect passengers)—with different insurance regimes associated with each. However, this approach places considerable responsibility on the user with the non-trivial risk that their selection of an egoistic mode results in the death of a pedestrian that might otherwise have been avoided. It also places responsibility for determining behaviour of the CAV in the three modes on the CAV developers, which again may not produce outcomes seen as socially or ethically acceptable. For these reasons, we oppose this approach.

Automated system manufacturers and deployers have a responsibility to reduce uncertainty and optimise behaviour (maximise utility) in the systems they develop. However, in a democracy, it is the responsibility of legislators as elected representatives to enact rules that govern our behaviour. They already do this regularly (e.g. UN, 1968). These rules reflect accepted community values and ethics, and in many cases include allowance for discretion or differential application to individual circumstances. If rules require an ethical goal function to reduce uncertainty (so they can be systematized or coded to enable them to be used by automated systems) then responsibility for stating that function should lie with legislators (who are democratically accountable to the public). Such a function must represent the will or need of society rather than that of CAV developers, as noted by Bonnefon et al. (2020), who state that their recommendations cannot be applied as a mechanical, top-down procedure, but rather need to be specified, discussed and redefined in context. They highlight the importance of inclusive deliberation to ensure that perspectives from all societal groups can be heard and no one is disregarded. Moreover, tensions between these principles may arise in specific applications. For these reasons, the design and development of CAV systems should be supportive of and resulting from inclusive deliberation processes involving relevant stakeholders and the wider public. We strongly support this approach and urge that this inclusive deliberation should be organised by governmental bodies and regulations [as is already done for example by the Global Forum for Road Traffic Safety (UNECE Working Party 1)]. This discussion was presented within Working Party 1 in the March 2021 meeting and received with great interest. Also, during the EU-CAD 2021 conference, our view that AI should primarily optimise on safety above other goal functions such as comfort or travel times was agreed upon by all those in the audience, measured via an online poll.

Once an ethical goal function has been agreed and enacted by legislators, CAV systems could use that function to determine the course of action with the highest utility. For example, how to minimise risks of harm while still enabling the CAV to travel from A to B within a specified set of parameters (e.g. date, time, route, type of vehicle, physical environment, applicable traffic rules and speed limits etc.). Applying this approach to whether or not to exceed a prescribed speed limit in a particular instance would require the CAV to calculate what speed would result in the highest utility in all of the other circumstances reflected in the data collected by the vehicle's sensors and any communication with third parties. Factors considered within the ethical goal function to determine the optimal choice of speed might include potential collision risk, suitability of the road, infrastructure, vehicle and weather and the presence and behaviour of other road users. A similar calculation that incorporates this function could assess whether mounting the pavement is permissible, in what circumstances and how that operation should be performed.

When considering what potential ethical goal functions need to be developed and how, the question arises whether they should apply continuously to all aspects of the vehicle's operation or whether they should only be applied in exceptional or high-risk situations. This could be considered analogous to prohibiting human drivers from allowing their vehicle to exceed the speed limit or mounting the pavement, except in situations where it is required as a reasonable response to an emergency or to maintain safety. However, adopting the exceptional or high-risk approach raises real problems in determining what conditions are deemed sufficient to warrant use of an ethical goal function as the basis for risk assessment by an automated system and how those situations could ever be sufficiently described—especially

given that they are by definition unusual, unexpected and unanticipated.

In a representative democracy, not every citizen or voter agrees with every provision of every piece of legislation. Nonetheless, for the democratic rule of law to be effective, citizens must agree that governing authorities have a social licence to impose policies and regulations. Although an ethical goal function for CAVs may not be universally accepted, it must have some form of democratic legitimacy. However, if incidents involving CAVs suggest that the vehicles have behaved in ways that do not conform with what the community expects or is willing to accept, legislators may find that they need to adjust the utility function or even consider withdrawing permission for CAVs to operate. Any ethical goal function for a CAV could therefore be subject to ongoing review based on observed performance in use. Importantly, and presumably prior to legislating that function, the outcomes of proposed ethical goal functions could also be explored and even experienced in simulations of CAV behaviour. Although some assumptions may be uncertain due to the unusual, unexpected and unanticipated of the situations to be encountered, rare events could even be simulated and would provide valuable insights into the definition of ethical goal functions for real vehicles (Aliman & Kester, 2019).

Although the critical role of the ethical goal function will only emerge in more extreme or rare situations, the first step forward in a democracy is inclusive deliberation. Public consultation regarding the implications of any function should be undertaken prior to any legislative action, to ensure that the function enacted reflects community values and notions of acceptable CAV behaviour. This is the crucial step in resolving the endless discussion of how a CAV should act in any imaginable situation. Studying whether understanding the role of an ethical goal function and its implication would make the public more or less receptive to the adoption of CAVs is important, since disagreements about the function may act as a disincentive to adoption and withhold any step forwards toward a solution. A continuous process of social and technological feedback (a so-called socio-technological feedback loop) is required to update and refine the goal function over time (Aliman et al., 2019). This would allow CAV behaviours to reflect changing societal norms, values and concerns. Confirming the essential role and accountability that governmental and legal bodies have in setting these rules is likely to provide assurance that CAV operations are governed at a high level by societal interests.

Vehicle manufacturers often seek to differentiate their products based on driving characteristics and capabilities. It is likely that manufacturers will seek to retain some differentiation between brands, even where automation can assume responsibility for some or all of the dynamic driving task. It may still be ethically acceptable for automated driving behaviours to differ markedly between vehicle types and brands so long as they use the same ethical goal functions. This presents an opportunity for further research into the extent to which consistent ethical goal functions would permit differentiation between vehicle types, and if so, what implications this would have for all stakeholders.

Ethical goal functions could allow for cultural differences between nations or regions. After inclusive deliberation, governments of different countries may formulate different ethical goal functions for CAVs (potentially varying over different domains of operation, e.g. in closed areas or within defined geographical limits). This clarity requires CAV manufacturers and deployers to apply the correct utility function within the relevant geographical area, even switching ethical goal functions in real-time for cross-border journeys. This would necessitate common descriptions, formats and protocols for the secure update of ethical goal functions in CAVs. It will also be critical to ensure that individual vehicle users cannot modify the ethical goal function of the CAV. This also provides clear boundaries for adaptive automation that automatically adjusts according to driver state or preferences. Automation can only adapt within the boundaries of the ethical goal function.

Conclusion

Current traffic rules have been created to help promote safe road use by humans and human-operated vehicles. In numerous ways, they allow for discretion and interpretation by human drivers based on a shared understanding of safety and behaviours in traffic, established through training, experience and common human characteristics (e.g. empathy, cooperation, patience etc.). Yet we have shown that there are profound disagreements between industry experts, road safety experts and public analysts over how CAVs should behave, even for relatively simple traffic rules. For CAVs to operate safely and ethically in shared public environments, they will need to display behaviours that show similar discretion and interpretation when necessary to maximise safety. However, given the infinite variety of driving scenarios that a CAV may encounter, these cannot be explicitly programmed. Conversely, deep learning approaches cannot establish ethical principles that underpin such decision-making.

We therefore conclude that to optimise road safety and resolve ambiguities in traffic rules for CAVs, their operation should be guided by ethical goal functions as part of a hybrid AI system. By allowing for an inclusive, public deliberative approach for the development of such functions, the safety behaviour of CAVs can reflect societal norms and values. This approach requires that citizens are informed about the implications of the ethical goal function and able to provide

input into the decision-making process of how this approach governs CAVs behaviour in critical situations.

We suggest that responsibility for creating the framework of CAV ethical goal functions should sit with an appropriate international body, for example within UNECE WP1. Specific values for the ethical goal functions within this framework can then be managed by relevant organisations in each country. In the UK, this could be the Department for Transport; in The Netherlands it could be the Ministry of Infrastructure and Water Management; in Germany, KBA (Kraftfahrt-Bundesamt); in Australia, DITRDC (Department of Infrastructure, Transport and Regional Development and Communications) and so on. The elected minister would be democratically accountable for defining what risks are acceptable in terms of road safety—just as they are for other aspects of road safety today. In this case, they would take the lead in determining the specific values for ethical goal functions that were most appropriate for their region. However, ethical goal functions would need to be developed through research and development first and validated in simulation with feedback from stakeholders, including the public, before being passed to the relevant department. Once established, the ethical goal functions need not stay the same over time—they can be improved via an ongoing collaborative feedback process in response to the observed safety behaviours of CAVs in the real world, with CAV developers updating their vehicles to reflect the new functions.

Defining ethical goal functions for CAVs enables developers to overcome uncertainties in how their vehicles should operate but does not require that regulators necessarily understand AI or utility functions. The standardised framework would enable vehicles travelling from one jurisdiction to another to update their ethical goal functions through a secure process to reflect the prevailing preferences for that region.

Our suggested approach has some alignment with the proposal of De Freitas et al. (2021), who set out five general recommendations for how CAVs can achieve driving 'common sense'—that they define using the acronym SPRUCE: driving behaviours should be safe, predictable, reasonable, uniform, comfortable and explainable. We agree that CAV driving behaviours will be defined by trade-offs, by operating under uncertainty, by the need to act based on available information and by the application of principles (rather than specified outcomes). We also agree that human driving may not represent the ideal reference behaviour on which to model CAV operations, that we need to understand behaviours in which responsibility for control may be shared between the human and automated systems, that training data will be insufficient for CAVs to learn how to drive successfully, that local customisation of CAV operating protocols will be important and that ethical CAV behaviours must emerge through consideration of various trade-offs across the transportation system. However, we believe 'common sense' can only emerge through an inclusive, deliberative government-led process that establishes a priori values for 'good' and 'bad' behaviours that can be used to define ethical goal functions against which developers can effectively and transparently optimise CAV behaviours, aligned to societal preferences.

Such functions also enable engineers to optimise CAV behaviour according to the maximum utility of the goal function and this can be developed in simulation, in controlled environments and in public road testing. In the event of a collision, a manufacturer can use the ethical goal function to demonstrate transparently how their CAV acted safely and ethically, according to the societally derived optimum. As evidence of how CAVs contribute to road safety emerges, the goal functions can be updated through continual socio-technological feedback loops. Whilst the ethical goal functions guide CAV safety behaviour, they do not specify other aspects of CAV operation, allowing manufacturers to differentiate their products according to brand preferences (e.g. comfort, sporty etc.)—provided they sit within the framework of the ethical goal functions.

Any ethical goal function must be developed through public consultation and deliberation, be democratically accepted and be communicated in such a way as to allow the public to have insight and influence over this process, with researchers and AI experts guiding this process. Without this, the community is likely to lack confidence in adopting CAVs, or CAVs will cause collisions or other adverse effects that could otherwise have been avoided. Cohesive action by regulators, developers and researchers will be required to formulate a utility function that adequately represents societal expectations regarding acceptable CAV behaviour and that can integrate successfully into the automated driving systems that are being developed.

There is much work to do in order to establish a practical approach to achieve this but our research suggests that this is an essential step in achieving the anticipated safety and efficiency benefits of CAVs and aligning different views and opinions on road safety. However, we have highlighted that such differences can be resolved through ethical goal functions that would give the public the opportunity to contribute to CAV behaviour, would give developers a solid basis on which to optimise CAV safety behaviours and would give manufacturers greater confidence in deploying CAVs that operate in line with both explicit traffic rules and implicit societal preferences. Ethical goal functions put society and human values back in control without the anticipated benefits of automation systems being lost in translation between lawyers, ethicists, citizens, road safety experts and AI experts.

Acknowledgements The authors would like to thank Dr. Nadisha-Marie Aliman for her helpful advice on various technical issues discussed in this paper and the individual academic and industry experts who engaged with our questions on CAV rule compliance, representatives from MIT, RWTH Aachen University and the Metropolitan Police for their insights in support of this paper and all people in our network that provided input to this paper with their ideas and opinions around behaviour of CAVs in case of breaking traffic rules.

Author contributions All authors contributed to the drafting of the manuscript. All authors have read and approved the final manuscript.

Funding No funding was received to assist with the preparation of this manuscript.

Data availability Submissions used from the Law Commission consultation on automated vehicles are publicly accessible (see references). Additional materials from interviews and correspondence with industry experts are not available due to commercial sensitivity.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Aliman, N. M. (2020). *Hybrid cognitive-affective strategies for AI safety*. PhD dissertation, Utrecht University, 2020. Retrieved April 2021 from: https://doi.org/10.33540/203

Aliman, N. M., & Kester, L. (2019). Extending socio-technological reality for ethics in artificial intelligent systems. In: *2019 IEEE international conference on artificial intelligence and virtual reality (AIVR)* (pp. 275–282).

Aliman, N.M., Kester, L., & Werkhoven, P., (2019). XR for augmented utilitarianism. In: 2019 *IEEE international conference on artificial intelligence and virtual reality (AIVR)* (pp. 283–285).

Aliman, N. M., Kester, L., Werkhoven, P., & Yampolskiy, R. (2019). Orthogonality-based disentanglement of responsibilities for ethical intelligent systems. In: *International conference on artificial general intelligence* (pp. 22–31). Springer.

Australian Bureau of Statistics. (2021) Criminal Courts, Australia. Retrieved June 2021 from: https://www.abs.gov.au/statistics/people/crime-and-justice/criminal-courts-australia/latest-release

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*. https://doi.org/10.2777/035239

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology & Work, 21*(1), 3–20.

Contissa, G., Lagioia, F., & Sartor, G. (2017). The ethical knob: Ethically-customisable automated vehicles and the law. *Artificial Intelligence and Law, 25*(3), 365–378.

Cummings, M. L., & Ryan, J. C. (2014). Shared authority concerns in automated driving applications. Retrieved April 2021 from: https://dspace.mit.edu/handle/1721.1/86937

De Freitas, J., Censi, A., Smith, B. W., Di Lillo, L., Anthony, S. E., & Frazzoli, E. (2021). From driverless dilemmas to more practical commonsense tests for automated vehicles. In: *Proceedings of the national academy of sciences*, 118(11).

Department for Transport (2020). SPE0112: Vehicle speed compliance in Great Britain. Retrieved February 2021 from: https://www.gov.uk/government/statistical-data-sets/vehicle-speed-compliance-statistics-data-tables-spe

Driver and Vehicle Standards Agency (2015). *The highway code*. TSO, ISBN: 9780115533426.

Dutch Safety Board (2019). Who is in control? Road safety and automation in road traffic. Retrieved April 2021 from: https://www.onderzoeksraad.nl/en/page/4729/who-is-in-control-road-safety-and-automation-in-road-traffic

Eckersley, P. (2018). Impossibility and uncertainty theorems in AI value alignment (or why your AGI should not have a utility function). Technical report, Partnership on AI & EFF. arXiv preprintarXiv:1901.00064

Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis & Prevention, 123*, 114–122.

European commission, proposal of the european parliament and of the council—Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union acts. 21 April 2021. Retrieved June 2021 from: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

European Parliament, European Parliament resolution of 20 October 2020 with recommendations to the commission on a civil liability regime for artificial intelligence (2020/2014(INL)). 20 October 2020. Retrieved 6 July 2021 from: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html#title1

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice, 77*, 167–181.

Goodall, N. (2021). Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2021.106056

Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M. H., Bengler, K., Andersson, J., Merat, N., & Reed, N. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, *20*(3), 223–249.

Law Commission (2018). Automated vehicles: A joint preliminary consultation paper. Retrieved February 2021 from: https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2018/11/6.5066_LC_AV-Consultation-Paper-5-November_061118_WEB-1.pdf

Law Commission (2019). Responses to the automated vehicles consultation 2018–19. Retrieved February 2021 from: https://www.lawcom.gov.uk/draft-responses-to-the-automated-vehicles-consultation-2018-19/

Legifrance (2021). Code de la route. Retrieved April 2021 from: https://www.legifrance.gouv.fr/codes/id/LEGITEXT000006074228/

Lu, Z., Happee, R., Cabrall, C., Kyriakidis, M., & de Winter, J. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 183–198. https://doi.org/10.1016/j.trf.2016.10.007

Marcus, G. (2020). The next decade in AI: four steps towards robust artificial intelligence. arXiv preprint. arXiv:2002.06177

Martens, M. H., & van den Beukel, A. P. (2013). The road to automated driving: Dual mode and human factors considerations. In: *16th international IEEE conference on intelligent transportation systems (ITSC 2013)* (pp. 2262–2267). IEEE.

Morando, M. M., Truong, L. T., & Vu, H. L. (2017). Investigating safety impacts of autonomous vehicles using traffic micro-simulation. In: *Australasian transport research forum* (pp. 1–6).

NHTSA (2017). *Office of Defects Investigation PE 16–007*. Retrieved April 2021 from: https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.pdf

NTSB (2018). *National transportation safety board preliminary report highway: HWY18MH010*. Retrieved April 2021 from: https://www.ntsb.gov/investigations/AccidentReports/Pages/HWY18MH010-prelim.aspx

NTSB (2019). *Collision between vehicle controlled by developmental automated driving system and Pedestrian Tempe, Arizona March 18, 2018*. Accident Report NTSB/HAR-19/03 PB2019–101402. Retrieved April 2021 from: https://data.ntsb.gov/Docket/Document/docBLOB?ID=40479021&FileExtension=.PDF&FileName=NTSB%20-%20Adopted%20Board%20Report%20HAR-19%2F03-Master.PDF

Prakken, H. (2017). On the problem of making autonomous vehicles conform to traffic law. *Artificial Intelligence and Law, 25*(3), 341–363.

Quimby, A. R., House, C., & Ride, N. M. (2005). Comparing UK and European drivers on speed and speeding issues: some results from SARTRE 3 survey. In: *Behavioural research in road safety*: *Fifteenth seminar*. London: Department for Transport (pp. 49–67).

Road Traffic Regulation Act 1984. Retrieved February 2021 from: https://www.legislation.gov.uk/ukpga/1984/27

Road Traffic Offenders Act 1988. Retrieved April 2021 from: https://www.legislation.gov.uk/ukpga/1988/53/contents

Road Traffic Act 1988. Retrieved March 2021 from: https://www.legislation.gov.uk/ukpga/1988/52/contents

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach*. Prentice Hall.

Rvv (1990) Traffic rules and traffic signs regulations. Retrieved April 2021 from: https://wetten.overheid.nl/BWBR0004825/2021-01-01

SAE (2018) J3016 Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Retrieved February 2021 from https://doi.org/10.4271/J3016_201806

Scanlon, J.M., Kusano, K.D., Daniel, T., Alderson, C., Ogle, A., Victor, T. (2021). Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. Waymo LLC. Retrieved March 2021 from: https://storage.googleapis.com/waymo-uploads/files/documents/Waymo-Simulated-Driving-Behavior-in-Reconstructed-Collisions.pdf

UN (1949). Geneva convention on road traffic. United Nations. Retrieved April 2021 from: https://treaties.un.org/pages/ViewDetailsV.aspx?src=TREATY&mtdsg_no=XI-B-1&chapter=11&Temp=mtdsg5&clang=_en

UN (1968). Vienna convention on road traffic. United Nations. Retrieved April 2021 from: https://treaties.un.org/pages/ViewDetailsIII.aspx?src=TREATY&mtdsg_no=XI-B-19&chapter=11

UN (2016). Report of the sixty-eighth session of the working party on road traffic safety. United National Economic and Social Council. Retrieved June 2021 from https://unece.org/DAM/trans/doc/2014/wp1/ECE-TRANS-WP1-145e.pdf

Wernaart, B. (2021). Developing a roadmap for the moral programming of smart technology. *Technology in Society, 64*, 101466. https://doi.org/10.1016/j.techsoc.2020.101466

WVW (1994) Road Traffic Act. Retrieved April 2021 from: https://www.hylaw.eu/database/national-legislation/the-netherlands/wegenverkeerswet-1994-road-traffic-act-912

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 285–307.