Driverless Futures?

Consultation response from the *Driverless Futures?* (driverless-futures.com) project team

Contact: Jack Stilgoe (j.stilgoe@ucl.ac.uk), Project principle investigator and lead author

February 2019

This is a response to the Law Commission's consultation on Automated Vehicles on behalf of the *Driverless Futures?* team (Dr Tom Cohen, Professor Peter Jones, Saba Mirza, Professor Graham Parkhurst, Dr Jack Stilgoe and Professor Alan Winfield). Our project, funded by the Economic and Social Research Council, is a three-year (Jan 2019 – Dec 2021) investigation of the governance options for self-driving cars. This project has not yet generated new data. Our consultation is therefore based on social science evidence and insight from the project team members, who have been involved with debates on transport, robot ethics and the governance of new technologies. Members of the team would be happy to discuss these issues further.

General comments

There is much to commend within the Law Commission's report, which represents one of the first serious policy attempts to engage with the short-term uncertainties and opportunities of AVs. The report clarifies some issues that are often intentionally or unintentionally blurred. The report uses the SAE classification, but there is additional consideration of the operational design domain for AVs, which is a limitation of the SAE scheme.

The report focusses on imminent governance questions, which is understandable. However, there are some assumptions made that limit the scope of the report. The first assumption, that innovation will take place on "Britain's existing road network", may hold in the very short term, but there are good reasons to suppose that some of the purported benefits of AVs will be realised with changes to infrastructure. AVs will, as the report identifies, be constrained by operational design domain. There will be pressure to make new domains machine-readable through improved lane markings, mapping and connectivity. The safety of future transport systems will depend as much on what is outside a car as inside. The report could usefully consider how the law may need to change in order to enable AVs through infrastructure improvements, but in ways that might impose costs, probably to the public sector.

UCL E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL UWE Bristol University of the West of England

Another questionable assumption is that there are two distinct 'development paths to full automation'. It is important to point out that there is no single trajectory of innovation. These two suggested paths will not necessarily end at the same point, and many of the benefits of the technology may be realised without 'full' automation. More precisely, as the report notes, automation will only be possible within particular constraints (defined by the operational design domain), meaning that 'full' automation will in reality always be conditional.

The testing of AVs on public roads will be necessary in order to train and certify software and hardware. Doing this responsibly and ensuring that the benefits and learning from testing are shared beyond just the AV developers will be vital. The recent framework for testing developed by the California DMV and Public Utilities Commission offers a useful alternative to the deregulatory approach favoured by other US states. European policymakers can and should go beyond California's approach to develop socially-robust, credible models of governance.

***Consultation Question 9** (Paragraphs 4.107 - 4.109):*
Do you agree that every automated driving system (ADS) should be backed by an entity (ADSE) which takes responsibility for the safety of the system?

The clarification of responsibilities through the designation of an automated driving system entity will help guide further policymaking. The proposal to create an assurance agency that is able to make sense of emerging AV technologies and approve them before market is therefore a good one. However, the capacity for technology assessment that this agency builds up should be used for more than just safety. As with health technology assessment bodies, the agency should also advise on the likely benefits of a particular technology, its wider implications and its performance compared to alternatives.

***Consultation Question 10** (Paragraphs 4.112 - 4.117):*
We seek views on how far should a new safety assurance system be based on accrediting the developers' own systems, and how far should it involve third party testing.

The use of independent testing will help improve public credibility and learning across different manufacturers. This could be particularly important for software systems where the incentives to make reasoning opaque are substantial.

***Consultation Question 11** (Paragraphs 4.118 - 4.122):*
We seek views on how the safety assurance scheme could best work with local agencies to ensure that is sensitive to local conditions.

The conditionality of AV systems will be a crucial focus for any safety assurance scheme, at least at an early stage, where systems are trained within, and geo-fenced by, limited, well-mapped environments. Local knowledge of infrastructure, patterns of road use, culture and

weather will all be determinants of safe operation. Safety that is assured in one place should not be seen as portable.

Although the question here focuses on the 'local', this concept should be extended to 'locale' in order to address the full implications of AVs being used across space. Goods vehicles and passenger cars registered in the UK will be used beyond UK borders, needing to cope with, among other things, variations in speed limits, quantitatively defined passing distances when overtaking, variable speed limits in wet conditions and more. The Irish border may become an early case-in-point here. The regulation of AV systems will need to be a global project.

***Consultation Question 13** (Paragraphs 5.54 - 5.55):*
Is there a need to provide drivers with additional training on advanced driver assistance systems?

Unlike with autopilot systems in aircraft, manufacturers have assumed that ADAS requires no additional driver training. The assumption that such technologies are clear and immediately usable is problematic. Evidence, including from studies in the UK funded by the UK government, of the 'handover' problem[1] reveals the difficulties that human drivers experience re-engaging with road vehicle control systems even after short periods of 'disconnection', with driving being measurably different for a period of minutes after retaking control. The NTSB's investigation of the May 2016 Tesla crash revealed a lack of clarity about the limits of the technology. Manufacturers and others should be encouraged to be clear about what a technology can't do, rather than exaggerate its capabilities. The PAVE campaign (pavecampaign.org) in the US is currently conflicted about its role. UK technology developers and policymakers can improve upon this.

***Consultation Question 14** (Paragraphs 5.58 - 5.71):*
We seek views on how accidents involving driving automation should be investigated. We seek views on whether an Accident Investigation Branch should investigate high profile accidents involving automated vehicles? Alternatively, should specialist expertise be provided to police forces.

The model of the US National Transportation Safety Board is worth considering. The NTSB is tasked with identifying the cause or probable cause of every air accident in the US, as well as highway crashes that they regard as sufficiently troublesome. The agency is separated from the courts system, which means that it can investigate crashes without consideration of liability, focussing on learning lessons and improving the safety of all technologies. If the focus is on improving safety, an accident investigation branch separate from the police

[1] Morgan, P., Alford, C., Williams, C., Parkhurst, G. and Pipe, A. G. (2017) Manual takeover and handover of a simulated fully autonomous vehicle within urban and extra urban settings. In: Stanton, N. A., ed. (2017) Advances in Human Aspects of Transportation: Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation. (597) Springer, pp. 760-771. ISBN 9783319604404 Available from: http://eprints.uwe.ac.uk/31248

would be advisable. The branch would also need to draw on specialist tools and knowledge. This may include software expertise in understanding the interpretability of machine learning systems. The use of the term "high profile" in the question is intriguing; cases should not be selected merely on their public prominence.

***Consultation Question 15** (Paragraphs 5.78 - 5.85):*
Do you agree that the new safety agency should monitor the accident rate of highly automated vehicles which drive themselves, compared with human drivers?
***Consultation Question 16** (Paragraphs 5.86 - 5.97):*
What are the challenges of comparing the accident rates of automated driving systems with that of human drivers?
Are existing sources of data sufficient to allow meaningful comparisons? Alternatively, are new obligations to report accidents needed?

We should not assume that the major consideration is the average accident rate. The publicly acceptable levels of risk are still unknown. The assumption behind the question is a consequentialist one, and will not be uncontentious. Members of the public will understandably disagree on how safe is safe enough, and they will ask 'safe enough for what?' An understanding of the risks of AVs cannot be cleanly separated from a perception of who benefits. AV crashes may be seen by the public in the same way as airline or train crashes are, where there is public outrage even if average death rates are two orders of magnitude less than in cars. The important prior question is whether a regulator should make requirements on data sharing and accident investigation so that safety is improved over time.

It is vital for regulators to clarify expectations for accident investigation before positions become polarised. The project of mandating event data recorders in conventional cars, which is still incomplete, provides a cautionary tale. Some manufacturers, including Tesla, have been very protective of their data, which they regard as proprietary. An accident investigation branch should determine what data is necessary for safety improvements and mandate its sharing in the event of crashes.[2]

Monitoring should go well beyond accidents and fatalities. Even though road fatalities no longer receive much public attention, fatalities from new technologies will. Fatal incidents, however, represent the tip of an iceberg of crashes and near misses that are all important for improving safety. A key metric for AVs is manual interventions, i.e. when the human driver has to intervene because an automated system either can't cope with a situation or fails completely. Evidence is hard to come by. The California DMV requires disengagement reports but lets developers decide where to draw the line. Some of these reports suggest disengagements are very common.

[2] Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. Social studies of science, 48(1), 25-56;

Manufacturers should have an obligation to report interventions relative to miles driven and to categorise them, e.g. minor interventions such as the car stopping or slowing unnecessarily or major interventions, without which there would have been a serious incident. Major interventions could require additional detail on the type of incident, speed of vehicles, other road users at risk. Without reliable data on disengagements and interventions it will be very difficult to assess safety, improve safety over time, build trustworthiness and inform changes to legislation.

***Consultation Question 38** (Paragraphs 9.6 - 9.27):*
We seek views on how regulators can best collaborate with developers to create road rules which are sufficiently determinate to be formulated in digital code.

It is vital that the process of defining and inscribing new road rules is an inclusive one. Policies here will not just 'govern the actions of highly automated vehicles'; they will govern the actions of others too. New rules may help clarify programming in the short term, but they could end up determining future behaviours and the shape of future infrastructures. Algorithmic rules should not become de facto rules of the road. Such rules could, if well-designed, make roads safer and fairer, protecting vulnerable road users and emergency vehicles. At the moment, many encounters on the road, especially in unusual scenarios such as those identified here, are governed by common sense. The formalisation of rules will place demands on other road users, who will be expected to know and respond to them. The negotiation of whether hard and fast rules are appropriate and how they should be written should be considered a democratic exercise, not a technical one. One need only look at recent debates around 'shared space' in urban design to see how contested apparently subtle shifts in norms and rules can become. Separate highway codes for machines and humans would create substantial complication for interactions on roads.

To give one example, certain rules within the highway code currently offer qualitative guidance, such as Rule 163, the safe passing distance for a motor vehicle overtaking a vulnerable road user. Such a distance will need to be quantified in an automated system, albeit with variable parameters according to circumstance. The process of determining the parameters will not be value-free nor objective, and needs to involve a wide consultation and likely further empirical evidence.[3]

***Consultation Question 43** (Paragraphs 9.68 - 9.74):*
To reduce the risk of bias in the behaviours of automated driving systems, should there be audits of datasets used to train automated driving systems?

[3] See the following for a review of some of the 'interactions' evidence gaps: Parkin, J., Clark, B., Clayton, W., Ricci, M. and Parkhurst, G. (2018) Interactions involving autonomous vehicles in the urban street environment: A research agenda. Proceedings of the Institution of Civil Engineers - Municipal Engineer, 171 (1). pp. 15-25. ISSN 0965-0903 Available from: http://eprints.uwe.ac.uk/33654

Recent research into algorithmic bias has revealed the flaws of a logic that suggests that human biases can be engineered out of a system with the use of algorithms. Biases can be introduced or exacerbated through data or through programming choices. Injustice can also arise from the development and use of tools that are designed, by well-meaning people, to tackle inequalities.[4] The acknowledgement of bias in large datasets has led to attempts by programmers to fix their algorithms. But such approaches fail to recognise that discrimination may be a feature of a system, rather than a bug. Algorithms are often used to classify things into categories. Even if they do so without fault, people may legitimately object to a particular classification. This is a political issue that cannot be 'fixed'. As well as questions of bias, regulators therefore need to ask about the purposes of particular algorithms and consider who benefits from their development and use.[5]

***Consultation Question 44** (Paragraphs 9.76 - 9.88):*
We seek views on whether there should be a requirement for developers to publish their ethics policies (including any value allocated to human lives)?

The German Ethics Commission on automated and connected driving drew a red line to prevent algorithms making choices between individuals in the event of unavoidable accidents. This is a useful principle, but it fits into a line of reasoning that outsources morality to machines rather than focuses attention on the responsibilities of system designers. As their report points out, thought experiments based on the so-called 'trolley problem' depend on a misrepresentation of the technology and act as a distraction from real governance questions. An example of an actual ethical choice made by designers was revealed by the March 2018 Uber crash in which Elaine Herzberg was killed. In this case, Uber engineers made a decision that was value-based as well as technical, to rebalance the system in favour of false positives. Regulators should focus on how AV systems can become safer over time rather than presuming that such systems are already intelligent enough to be able to make moral calculations in the event of crashes.

***Consultation Question 46** (Paragraphs 9.91 - 9.93):*
Is there any other issue within our terms of reference which we should be considering in the course of this review?

An important issue relates to the labelling of Automated Vehicles. Some developers have claimed that, if vehicles are labelled, other road users may behave differently in their presence, stopping them from functioning properly. The debate around agricultural biotechnology suggests that the labelling of the technology could be a politically contested area. A principle worth bearing in mind is Toby Walsh's so-called Turing Red Flag law ("An

[4] Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
[5] For more on this, see Kate Crawford, 2017, The Trouble with Bias - NIPS 2017 Keynote https://www.youtube.com/watch?v=fMym_BKWQzk

autonomous system should be designed so that it is unlikely to be mistaken for anything besides an autonomous system, and should identify itself at the start of any interaction with another agent.")[6] This point is also captured in one of the EPSRC's principles for robotics ("Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.")[7]

[6] Walsh, T. (2015). Turing's red flag. arXiv preprint arXiv:1510.09033.
[7] https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

Driverless Futures?

About the project (www.driverless-futures.com)

Self-driving cars or automated vehicles (AVs) promise to be one of the most disruptive technologies of the 21st Century. Proponents imagine them as a solution to problems as varied as road safety, sustainability, traffic, and accessibility. Governments in the UK and elsewhere see the potential to secure economic growth and new high-tech jobs. The UK's Industrial Strategy, launched in 2016 and refreshed in November 2017, has AVs as a priority, backed by substantial investments in science and engineering. Consultants and investors are optimistic. Morgan Stanley (2014) forecasts a multi-trillion dollar global market with billions of extra dollars in productivity gains in a 'New Auto Industry Paradigm'. KPMG (2012) calls AVs 'The Next Revolution'.

The history of science and technology tells us that paradigm shifts and industrial revolutions are rarely smooth and never just about science and technology. Policy reports have already identified the potential for self-driving cars to worsen inequalities by taking away millions of driving jobs. The development of the technology is not pre-ordained, nor will it be unproblematic. As self-driving cars encounter pedestrians and other vehicles on the road, new questions of responsibility come to the fore.

We should ask who is driving innovation. Are we ready to hand over control of our cars and our futures to the developers of self-driving cars? Can we anticipate the politics of self-driving cars? How should we imagine alternative futures? What do members of the public think? How should new technologies be governed?

The Driverless Futures? project is led by Dr Jack Stilgoe. It starts in January 2019 and ends in December 2021. It is a collaboration between University College London and UWE Bristol, funded by the UK Economic and Social Research Council. We will be working with the people developing the technology as well as with public groups, and we will establish a hub for international collaboration and comparison.

UCL

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

UWE Bristol University of the West of England