July 2023

SELF-DRIVING BUT GUIDED BY PEOPLE

How to make automated vehicles ethical

Project funded by

Rees Jeffreys Road Fund

reed mobility
Wokingham, UK
Nick Reed
nick@reed-mobility.co.uk

dg:cities
London, UK
Ed Houghton
ed.houghton@dgcities.com

TRL
Wokingham, UK
Ashley Pressley
apressley@trl.co.uk

HUMANISING AUTONOMY
London, UK
Kim Vigilia
kim@humanisingautonomy.com

april6
Bournemouth, UK
David Dent
david.dent@aprilsix.com

CONTENTS

FOREWORD

This project was funded through a competition organised by the Rees Jeffreys Road Fund to celebrate 150 years since the birth of its namesake, William Rees Jeffreys (1871-1954). Entrants were asked to imagine their vision of:

"...the way in which our roads (motorways, highways or streets) could best work for us all as we square up to the challenges of the next 50 years?"

This reflects the original vision of Rees Jeffreys who campaigned for more attractive, safer and more accessible roads and streets – and that providing such infrastructure would encourage people to get out and about and enjoy the UK's towns, villages and countryside.

It is widely anticipated that self-driving vehicles will play an increasingly significant role in our transport system over the coming decades with the much-vaunted promise that they will greatly increase the safety, efficiency and accessibility of our roads. However, as technology developers rush to build, test and trial such vehicles, it seems an important element is being overlooked – making sure that self-driving vehicles are deployed and operated in such a way that the communities affected by their presence come to appreciate them.

In response to the challenge set by the competition, the proposal for the project described in this report sought to create a process that would enable technology developers and transport regulators to engage with the public on self-driving vehicle technology. In doing so, we might begin to understand how the operation of such vehicles should be optimised to align with their preferences. As a result, the intention is that road use by future self-driving vehicles will truly deliver the best outcomes for us all and thereby fulfil the vision of William Rees Jeffreys.

I am delighted to be able to share the results of this important project with you and want to express my sincere gratitude to the Rees Jeffreys Road Fund for choosing to support this project and to my project partners for their superb support in its successful delivery.

Dr Nick Reed

Founder, Reed Mobility

Figure 1. William Rees Jeffreys (1871-1954; picture from Rees Jeffreys Road Fund)

EXECUTIVE SUMMARY

Car culture has been a defining feature of transport policy, urban design and individual mobility decisions over the last century. Cars, buses, trucks and motorbikes, the construction of roads upon which they drive and the development of supporting infrastructure (for fuelling, maintenance, regulation etc.) have enabled personal, societal, commercial and industrial transportation – and thereby facilitated **connectivity** and **prosperity**. This has come at a price. We now recognise the environmental harms associated with motorised road transport and millions have died around the world as drivers, occupants, riders or pedestrians in road crashes.

For almost as long as cars have been available, the idea of automatic vehicle control has been imagined. Mirroring autopilot in aircraft, futurists of the early twentieth century imagined vehicles joining electronically controlled highways – at which point, the driver could relax and allow the system to operate their vehicle safely and effectively towards their chosen destination. With developments in hardware, software and sensors, **self-driving vehicles** emerged as an achievable proposition in the early years of the 21st century. This prospect fuelled hype around the potential transformation in mobility that such vehicles might deliver, promising to improve **safety**, **efficiency** and **accessibility** of transport – and thereby reduce some of the negative externalities that are associated with car culture. Vehicle manufacturers, technology companies, research organisations and start-ups all clamoured to claim the latest breakthrough or innovation, prompting investors to pour funding into the sector.

Optimistic predictions made by some in the early 2010s for the subsequent rapid proliferation of self-driving vehicles in the years that followed have not been realised. Only one or two genuine self-driving vehicle-based transport services operating on public roads have emerged in the early 2020s. One of the challenges in developing the technology has been how to enable self-driving vehicles to manage the **infinite variety** and **complexity** of the scenarios they face in real world driving safely and reliably. A common approach to tackle this issue is to use **machine learning** – giving an automated driving system repeated exposure to road scenarios and allowing the system to associate its perceptions and actions with the correct driving behaviours. Whilst this has similarities to the way humans learn to drive, human drivers are given explicit instructions about how to operate a vehicle and the driving skills that emerge sit alongside a wide range of **learned** and **evolved perceptual** and **social skills** that we apply to understand the world around us and determine appropriate behaviours.

Reed et al. (2021) proposed that to address this mismatch, artificially intelligent systems that govern the behaviour self-driving vehicles should be required to follow Aliman and Kester's (2019) concept for '**ethical goal functions**'. These are a mathematical description of society's expectations over the behaviours considered to be desirable – and it was suggested that these should have 'democratic legitimacy'. This means that the communities affected by the deployment of the technology should have some input into the guidance of its behaviour.

Reed et al. (2021) did not elaborate on how communities could have such an input. This project, funded by the **Rees Jeffreys Road Fund**, has attempted to do that. Led by **Reed Mobility** and working with **DG Cities**, **TRL**, **April Six** and **Humanising Autonomy**, a survey of more than 2,000 participants and two workshops were conducted to investigate societal preferences for the behaviour of an urban self-driving bus service. This was supported by an advisory panel comprising experts in self-driving vehicles, computer science, social research, safety assurance, human factors and artificial intelligence.

Our work proposed eight '**dimensions**' of ethical value associated with self-driving vehicles to explore which the participants would prioritise. This was examined through the lens of survey responses and a specific ranking exercise in the surveys. Three dimensions seemed to feature most prominently. They were **safety** (of self-driving vehicle passengers and other road users), **legality** and **trust**. It is suggested that of those three, trust is the overarching priority since participants would build trust from knowing the vehicle operates safely and within regulations defined by a competent authority.

By contrast, the dimension given the lowest priority was cost – participants were clear that safety should not be compromised in the interests of cost saving.

EXECUTIVE SUMMARY

Building on the theme of trust, the workshops and surveys identified several '**ethical red lines**' – key facets of self-driving vehicle behaviour for which there seemed to be broad support – and which could therefore be considered as a starting point for the EGF:

- The deployment of self-driving vehicles should improve road safety.
- They should operate within a clear legal framework.
- They should not take risks to save time or to reduce cost.
- All road users should be protected equally.
- Data sharing with reasonable stakeholders (e.g. insurers, police etc.) is acceptable, provided this is done transparently and without conflicting with data privacy and data protection regulations.

As has been observed in many other studies of societal acceptance of technologies, positivity towards self-driving vehicles decreased with age but increased for the eldest category of participants. It is suggested that respondents in this category recognised the potential benefit of self-driving vehicles in supporting their independent mobility into older age.

Workshop participants strongly suggested that self-driving vehicles must always obey road rules but when faced with a choice between causing an injury collision and contravening a road rule, survey respondents suggested that a self-driving vehicle should **minimise harm** rather than comply with the rule. It is suggested that this divergence should be explored in further work. Some differences between participants' responses may have been a consequence of variation in their understanding of the described scenario, the capabilities of the technology and the outcomes that result from its actions.

We propose that **simulation** studies in which participants **observe** the behaviours of self-driving vehicles in tightly controlled and choreographed scenarios and then report the acceptability of those behaviours could generate useful data to help define the parameters of the ethical goal functions. The results would help **developers** to understand how their SDVs should behave, **regulators** to understand how they should moderate SDV behaviour and the **public** would have greater confidence that SDVs will operate in line with their expectations.

In their commentary on this project, the technology developer, **Humanising Autonomy**, recognised the limitation of black box approaches in safety critical applications, echoing the assertion that such systems cannot derive ethical values – and that these must therefore be developed explicitly. Our work has explored techniques by which these might be derived for self-driving vehicles in line with the recommendations of Reed et al. (2021).

This project has begun to elucidate a process and structure by which regulators can engage with the public in the acceptable and desirable features of SDVs. It is an approach that is not just applicable to SDV technology but also to other AI-based transport innovations that may emerge – and to help maximise the probability they are welcomed and embraced by an overwhelming majority and not scuppered by a minority of vocal opponents. In the coming fifty years, we believe that this will be vital if we are to fulfil the vision of Rees Jeffreys and ensure that our roads genuinely deliver prosperity and enjoyment for future generations.

INTRODUCTION

The emergence of car culture

Many towns in Britain can trace their origins back thousands of years to the Stone Age. When cars arrived in the late 19th and early 20th century, many roads and streets that exist today had already been established. As cars, trucks, buses and motorbikes became more accessible to more people, their value in delivering mobility of people, goods and services was obvious. With the growth of motorised road transport, particularly after World War II, the focus of planning in some locations shifted towards accommodating motor vehicles with wider paved roads and parking infrastructure. For example, (see Figure 2) shows a wide, three lane carriageway on the approach to a 14th century church in Coventry – a city which became known as the UK's 'Motor City' due to the growth of the car industry in that region. The U.S. experienced rapid growth in tandem with the rise in the availability and affordability of cars. As a result, the infrastructure of its cities and neighbourhoods is characterised by car usage and ownership to an even greater extent – exemplified by the like of Detroit and Los Angeles with central grid layouts and wide highways that criss-cross the city (see Figure 3).

The emergence of this emphasis on roads for transportation did not happen by accident. Many decisions have contributed to creating our infrastructure in this way. These include transport, industrial and residential planning, subsidies for automotive and energy companies, advertising expenditure and individual lifestyle choices. These decisions have enabled convenient access to independent mobility for individuals and businesses, supporting prosperity in many forms.

However, we now appreciate some of the problems associated with car dependence, including delays caused by congestion, poor air quality from exhausts, tyres and recirculation of particulates and deaths and serious injuries from road crashes. Furthermore, we recognise the contribution that road transport can make in causing climate breakdown (Department for Transport, 2022), poor health through inactivity (Hickman, 2019) and reinforcing inequalities across our communities (Mattioli, 2021).

It is possible that some of these negative consequences may have been less pronounced if the communities affected by these decisions had a better understanding of their potential impact and had more of an opportunity to influence their outcome. Community engagement is now recognised as a critical element in the transport planning process with local authorities using digital platforms (in addition to traditional engagement techniques) to enable citizens to raise issues about existing infrastructure or to provide feedback on proposed schemes. However, as we progress towards a future in which self-driving vehicles[5] (see Figure 4) are expected to play a significant role in the movement of people, goods and services, there has been a lack of public engagement on this topic.

[5] *For the avoidance of doubt, this report uses the phrase 'self-driving' to refer to vehicles that use a variety of technologies to deliver the full function of the dynamic driving task without the need for monitoring or input from a human driver (BSI, 2022).*

Figure 2. Three lane road plus bus stop (Queen Victoria Road) on approach to St. John the Baptist church (founded 1344) in Coventry, UK.

Figure 3. Judge Harry Pregerson Interchange, Los Angeles, U.S.A.

INTRODUCTION

In some ways, this is understandable – given the number of people killed and seriously injured in road crashes each year, it is reasonable to assume that people would welcome the arrival of technologies that could help to reduce the tragic toll of casualties. However, these anticipated (but as yet unproven) safety benefits wrapped in the gloss of a captivatingly futuristic innovation may conceal some of the less positive aspects of the technology. For example, self-driving vehicles will:

- ...be subject to system wide risks where all vehicles with the same hardware / software would be exposed to the same error or fault;
- ...behave in unexpected ways due to the infinite variety of (traffic, weather, road etc.) conditions that they will encounter when travelling in a public environment;
- ...bring changes in employment for those currently working as professional drivers.

Such topics have rarely been at the forefront of discussions about the development, regulation and deployment of self-driving vehicles. Learning from the emergence of car culture, the opportunity exists to engage more effectively with society in shaping the roles and behaviours of self-driving vehicles. This engagement should help to align behaviours of self-driving vehicles with the expectations of those who use or encounter them and ultimately ensure that they enhance the communities into which they are will be deployed.

The arrival of self-driving vehicles

Building on research funded by a U.S. defence agency, Google announced in 2010 that it was working on a self-driving car programme (see Figure 5). Envisaging that this technology might prompt a fundamental transformation of the way we use our roads, the announcement triggered a wave of activity across the technology, automotive, insurance and regulatory sectors to develop the required systems and legislation to enable automated driving.

Self-driving vehicles (SDVs) have been promoted by automotive companies, technology developers and governments as a potential route safer and more efficient transportation (e.g. Waymo, 2018; Cruise, 2014; HM Government, 2022; Coalition For Future Mobility, 2017). Their intuitive appeal is compelling. However, whilst a few small-scale commercial services using SDVs have started to appear in defined locations (e.g. Waymo, Cruise), their emergence as a significant contributor to the transport system has been far slower than was anticipated by many of their proponents. For example, in 2016, a Medium post by Lyft's co-founder, John Zimmer, envisaged that by 2022, around 80% of all their ride-hailing trips would be delivered by SDVs (see Figure 6) and that car ownership would end in major U.S. cities by 2025 (Zimmer, 2016).

Figure 4. Example of a self-driving vehicle: an adapted Nissan Leaf being trialled in London as part of the ServCity project (2018-2023).

Figure 5. Testing an early version of Google's self-driving car in 2011 (Image credit: Steve Jurvetson, CC BY-SA 2.0)

INTRODUCTION

Three interlinked factors seem to have caused the disparity between expected and actual operation of SDVs. The first factor is **technological**. Challenging though it is, creating a robotic vehicle that can perceive its environment, plan a route from origin to destination and then accelerate, brake and steer appropriately along its chosen route is the easy part. Building systems capable of doing that whilst also understanding the complexity of driving in mixed traffic, making valid predictions about the future behaviours of other road users and successfully driving day-after-day in a variety of light and weather conditions is a far greater challenge.

The second factor is the **regulations** that govern SDVs. National and international rules that define roads, vehicles, certification, driving behaviours and insurance requirements have all been established over more than a century of operation of human-driven motor vehicles. The emergence of SDVs has provoked a broad reappraisal of these regulations (e.g. Law Commissions, 2022[6]). The third factor is **public acceptance**. This phrase somewhat misrepresents the issue suggesting that people learning to tolerate SDVs is a particular barrier that must be overcome and is somehow separate from the development of the technology itself (Stilgoe & Cohen, 2021).

However, provided the potential benefits are genuine, there is an important task to demonstrate to the public the anticipated benefits that SDVs could bring to their lives, livelihoods and communities and the guard rails that are in place to militate against the perceived downsides (such as concerns over a loss of freedom and independence, safety risks, job losses etc.).

Ethics and self-driving vehicles

As we await widespread deployment of SDVs, one topic that has gained significant attention is the ethics of their operation, with ‘The Moral Machine experiment’ (Awad et al., 2018) attracting worldwide interest. Adapting the ‘trolley problem’ thought experiment (Foot, 1967), this paper describes a study in which hundreds of thousands of participants from around the world were asked to select their preferred behaviour of a self-driving vehicle in response to abstract, fictional moral dilemma situations – for example:

“An automated vehicle experiences a sudden brake failure. Staying on course would result in the death of two elderly men and an elderly woman, crossing on a “do not cross” signal. Swerving would result in the death of three passengers, an adult man, an adult woman, and a boy.”

Example moral dilemma from the The Moral Machine Experiment (Awad et al., 2018).

The study was an exploration of how participants’ preferences could contribute to developing global, socially acceptable principles for machine ethics. It produced fascinating results showing cross-cultural variation in the underpinnings of ethical decision-making. However, it has been criticised for narrowing the debate around SDV ethics to imaginary (or at least extremely unlikely) dilemma situations with the upshot that wider ethical concerns are overlooked (e.g. Etienne, 2019).

The European Commission commissioned an expert group to produce a report that sought to address these wider issues (Bonnefon et al., 2020).

This provided twenty recommendations addressing the ethics of SDVs, covering topics such as road safety, privacy, fairness, explainability and responsibility. With respect to road safety, the report suggested that SDVs should decrease harm experienced when compared to that caused by 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).

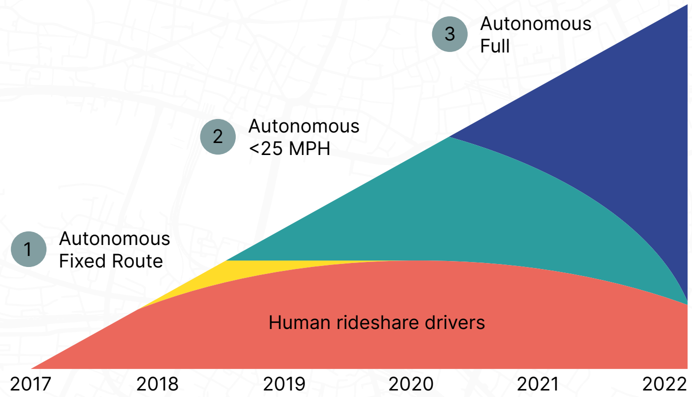

Figure 6. In hindsight, an overly optimistic prediction for self-driving vehicle adoption by Lyft for ride-hailing services from a 2016 Medium post by its co-founder, John Zimmer (Medium, 2016)

[6]*‘Law Commissions’ is used to refer to the joint activity undertaken by the Law Commission of England and Wales and the Scottish Law Commission in reviewing the UK regulatory framework for automated vehicles click here.*

Compliance with traffic rules

Traffic rules have been established to help guide safe, consistent and predictable behaviour by and for road users. However, they do not guarantee it. In some situations, departure from strict compliance may be required to minimise harm. This is reflected in traffic rules by 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, motivations, habits, social norms, and a general understanding of the road environment and the behaviours of other road users.

Programming SDVs to exhibit similar discretionary behaviours in the interests of road safety is challenging as it would require developers either to program how SDVs should behave in all foreseeable situations or train the artificial intelligence (AI) systems in how to behave in such situations. Reed et al. (2021) suggested enabling AI systems to deduce correct behaviours through exposure to a large number of training cases requires three extremely challenging practical difficulties to be overcome:

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 SDV 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.

Ethical goal functions

To address the challenge of ensuring that SDVs behave in an ethical manner, Reed et al. (2021) suggested an alternative approach. They proposed that SDV behaviours should be guided by 'ethical goal functions' (EGFs). This term originates in artificial intelligence research (Aliman & Kester, 2019) and suggests the creation of a mathematical description of the societal values that should underpin the behaviours of a complex system.

Reed at al. (2021) proposed that EGFs could be tailored geographically – the parameters of the EGF should reflect the locations in which the SDV was designed to operate and the expectations and behaviours of the communities with which it interacts. The EGF should evolve – as societal expectations change, the EGF should be updated accordingly. The EGF should also permit variation in behaviour – for example, a manufacturer could design (or users might choose) for their vehicles to offer more 'sporty' or more 'comfortable' driving modes. This is acceptable provided the behaviour of vehicles in each mode remained within the parameters of the EGF.

Importantly, Reed et al. (2021) suggested that the development of EGFs for SDVs should have '**democratic legitimacy**' – the communities affected by the deployment of SDVs should have the opportunity to contribute to the function that governs their behaviour and that the process should be overseen by an appropriate government agency so that those unhappy with SDV behaviour and EGFs can express their dissatisfaction when casting their vote.

As discussed above and as noted by Aliman and Kester (2019) an EGF cannot be 'learned' by the automated system. It must be constructed and defined by humans. However, whilst Reed at al. (2021) explored the application of EGFs for SDVs, they did not suggest a process by which the preferences of a community could be captured and translated into an EGF. This research represents a first attempt at doing so – with the aim that the approach supports both developers and regulators of SDVs in creating and regulating technologies that are aligned to the desires and expectations of the communities into which they are deployed.

We use a survey and workshop approaches to explore how the public expect SDVs to **behave**, how they should be **operated** and how they should be **regulated**. We attempted to identify '**ethical red lines**' – critical principles related to SDV operation for which broad agreement emerged. We then sought commentary from expert technology developers (**Humanising Autonomy**) on how those principles might be integrated into SDV workflows. We also discuss the wider application of this approach as a response to the growing use of technology in the mobility sector.

METHODOLOGY

This study deploys a mixed methods approach to provide insight into public perceptions of the ethics of SDV operation. We conducted a quantitative online survey across the UK to collect data from a broad demographic to identify trends, patterns, and common themes.

For richer insights into perceptions of safety, we complemented the survey with a qualitative methodology utilising two in-person deliberative workshops. Deliberative workshops were selected as they are deemed suitable supporting consensus building for topics where there could be variance in people's opinions.

The approach also encourages active discussion and learning among participants. The workshop format involves sharing information in phases and giving participants the opportunity to learn more about a topic, consider relevant evidence and discuss this evidence before presenting their view.

Self-driving vehicle application

Within the constraints of the project, it was considered impractical to try to cover all the ethical factors associated with all potential applications of SDVs (robotaxis, goods deliveries, shuttle buses etc.). In both the survey and workshop setting we therefore used the example of a self-driving shuttle bus operating in an urban environment. The information provided to participants about the use case is provide in Appendix A. This was chosen as one of the likely candidates for early public deployment of SDVs and to align with the expertise of project partners (Smart Mobility Living Lab and DG Cities).

Survey

The survey was distributed online between the 10th November and 1st December 2022. Respondents were predominantly recruited via Facebook and a smaller number completed the survey from a link posted on LinkedIn. The survey comprised basic demographic and transport use questions and 30 statements relating to SDV operation (listed in Appendix B). Participants were asked to indicate the extent to which they agreed with each of these statements using a 0-10 scale (Completely disagree-Completely agree). The median completion time was just over 11 minutes. Participants were incentivised by entry into a prize draw. After data cleaning, we analysed 1,502 good quality responses.

For 28 of the 30 statements, a positive or negative association was assigned with one or more of eight factors:

- **Mobility**: The ability for you and people in your community to travel easily to work, school, hospital, to see friends etc.
- **Legality**: Self-driving vehicles must obey the rules of the road at all times
- **Trust**: Trust that self-driving vehicle manufacturers and government regulators have done everything necessary to ensure that self-driving vehicles are safe and ready for the road.
- **Safety of passengers**: Operating a self-driving vehicle to protect the safety of passengers within the self-driving vehicle.
- **Safety of others**: Operating a self-driving vehicle to protect the safety of other road users such as pedestrians and cyclists.
- **Cost**: The full cost of delivering the self-driving vehicle service, including costs of development, cost of keeping safety data records, insurance costs, value of IP (remembering that higher costs for development and operation would likely mean higher ticket prices and/or less frequent services)
- **Fairness**: Self-driving vehicles operating in a manner that treats road users fairly (e.g. giving more space to pedestrians than to other vehicles)
- **Urban design**: The extent to which we need to adapt our streets to accommodate self-driving vehicles e.g. their own lanes and signage, barriers to prevent pedestrians crossing in front of them etc.

As an example, participants were asked to rate their agreement with the statement:

"I would want the self-driving bus to take emergency action to avoid a pedestrian who unexpectedly stepped in the road, even if it meant risking injury to passengers on board the vehicle."

Their rating of agreement with this statement was given a **positive** weighting towards the factor 'Safety of others' but a **negative** weighting (i.e. the rating value was multiplied by -1) towards 'Safety of passengers'.

The assignment of weightings for all statements in the survey is shown in Appendix C. Participants' agreement with each statement were summed and averaged, applying the positive and negative weightings for each factor, to give an estimate of the importance that participants placed on each factor. The key outcomes from the survey were used to inform the design of the subsequent workshops.

Participants were incentivised to participate in the survey by entry into a prize draw with a £100 voucher and five £30 vouchers awarded to participants, selected at random from all those who supplied details for entry into the draw.

Workshops

Two facilitated in-person workshops were held in London at the Smart Mobility Living Lab. Participants were guided through a series of exercises designed to support and facilitate dialogue and debate. Participants were recruited via an option to take part from the survey instrument and via adverts placed on Facebook. Each workshop lasted for approximately three hours and participants were given a £50 voucher as compensation for their time and expenses for participation.

A total of fifteen people took part (nine participants who had accepted the invitation to participate did not attend) across the two workshops. All participants were over 18 years old but ages were broadly distributed across the eight male and seven female participants. Two facilitators ran a series of exercises to enable discussion, and captured notes on the MIRO workshop facilitation platform. Exercises were developed based on the results from the survey that were of particular interest and included:

- Sentiment mapping: Undertaken at the start and end of the session to map changes in perception of self-driving services.
- Exercise 1: Risk taking behaviours scenario.
- Exercise 2: Road safety scenario.
- Exercise 3: Data privacy and data sharing scenario.
- Exercise 4: Dimension ranking exercise.

Figure 7. Workshop activities being led by DG Cities at the Smart Mobility Living Lab, London.

Advisory panel

This study has been supported by an advisory panel, chaired by a world-leading computer scientist working in the field of safety engineering and comprising twelve experts from the public, private and academic sectors. The advisory panel met three times – firstly, to confirm the suitability of the study design; secondly, to review the results from the survey and thirdly, to review this report prior to publication. All members supported the advisory panel as individuals rather than as representatives of any specific organisation. The authors are very happy to acknowledge their superb input to the project.

RESULTS

The key findings from both research activities are presented below. The next section provides an overview of the results from the survey followed by a summary of the outcomes from each workshop exercise. In each section we outline the 'ethical red lines'. These red lines represent our interpretation of where broad consensus was reached on each issue under discussion.

Key results:

- While some statements showed relatively consistent views between participants, many showed significant diversity of opinion, highlighting the challenge in defining features of an ethical goal functions that would be universally acceptable to a community.
- Participants become increasingly less positive towards SDVs with age until they reach 65 years old when they start to become more positive, perhaps with a view to the potential for SDVs to support independent mobility into older age.
- Trust, legality and safety (of other road users and passengers) appear to be the most important features of the operation of SDVs.
- In line with the recommendations made by the European Commission report (Bonnefon et al., 2020), SDVs must be no less safe than human driven road vehicles performing equivalent journeys.
- Safety should not be sacrificed in the interests of cost or time saving and when incidents do occur involving SDVs, we must learn from them to prevent them happening in future.
- Transparency and data sharing (with government / an appointed authority / insurance providers) to support improved safety, accountability and service was seen as very important but data protection and privacy issues must be addressed in doing so.
- However, while participants concede that it is difficult to eliminate risk from the transport system (noting that this might often arise from the actions of other (non-automated) road users), individuals indicated differing levels of acceptance of residual risk.
- SDVs should be compliant with rules by default, even if non-compliance would save time or aid the transit of an emergency vehicle.
- Participants indicated that safety performance should be standardised but with the possibility that operators could tailor driving styles to suit the environment in which their SDVs were driving.
- A high level legal and governance framework was deemed important by participants, not only for the purposes of safe operations, but also to ensure that users are able to trust service providers and vehicle manufacturers.
- The safety of other road users was considered to be at least as important as those in and SDV (perhaps more important as other road users may not have explicitly chosen to engage with the technology whereas an SDV passenger had made a positive decision to use one).

Survey insights

In this section, we summarise the findings from the survey, drawing out key insights from the study.
Results are presented by theme.

Trust

This theme in the survey explores perceptions of trust in the systems intended to govern and provide self-driving bus services. Survey respondents were asked to consider the extent to which they trusted different actors, including government, technology developers and service providers:

It can be seen that trust towards technology developers (Figure 8) or the government (Figure 9) diminishes with age and reaches its lowest for the 55-64 year old group with a slight uptick for older participants. Similar results can be observed across many of the statements regarding positivity towards SDVs. We suggest that, in the eldest age groups, there is greater support for SDVs in recognition of the perceived usefulness (Davis, 1985) for such technology to support independent mobility in later life.

I would trust technology compaines (eg. Google, Apple) to produce self-driving buses that operate safely on UK roads, by mean

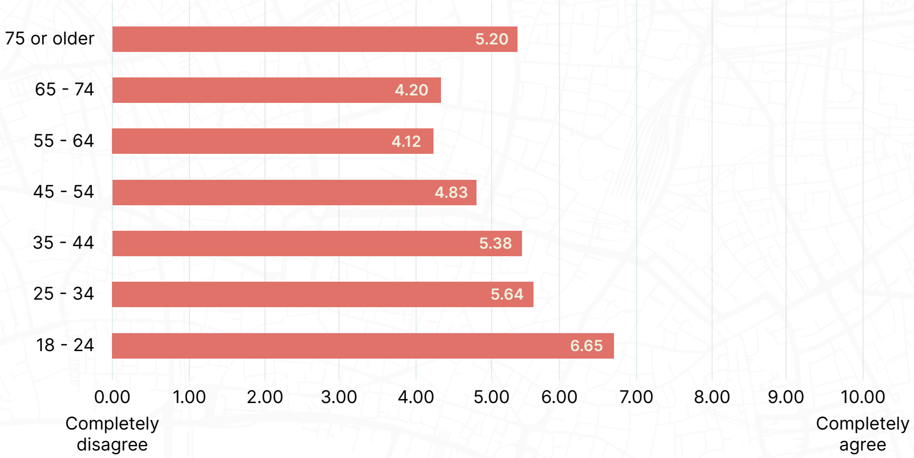

Figure 8. Trust in technology developers (mean response by age).

The UK goverment should not be heavy-handed over safety regulations that might delay the growth of the industry, by mean

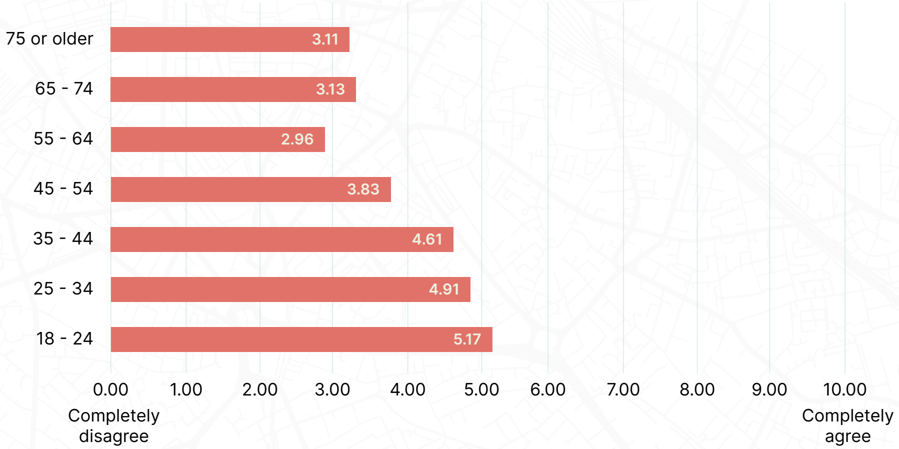

Figure 9. Trust in the government and regulations (by age).

Data

This theme in the survey explores perceptions of data privacy, data-sharing and data management by providers of self-driving bus services. Survey respondents were asked various factors related to data to measure their views on the safe and appropriate use of personal and aggregate data.

Views on data sharing and data utilisation by service providers were some of the most consistent across all survey participants. Overall, there was significant positive agreement that data sharing was important following an incident to determine blame (Figure 10).

The data collected by a self-driving bus must be shared with government investigators in the event of a crash to help understand why it happened and who was to blame.

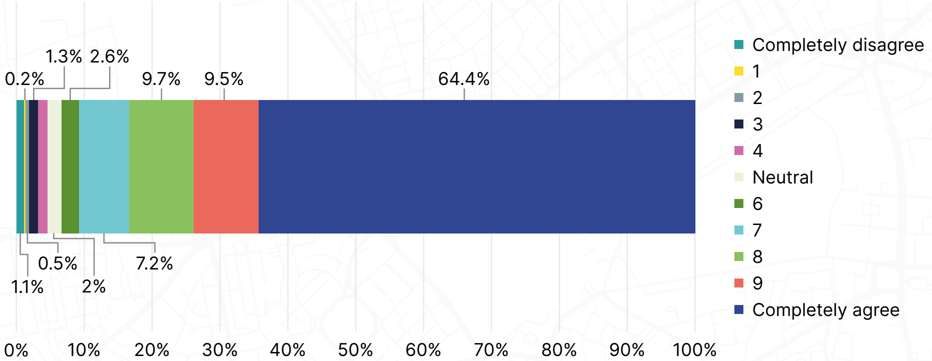

Figure 10. Data sharing in the event of an incident (mean = 8.94)

Risk

This theme in the survey explored perceptions of risk taking by the self-driving bus and was used to determine public sentiment towards risky road manoeuvres. Responses under this theme highlighted greater distribution of views, whilst some were comfortable with higher risk manoeuvres, the largest proportion were more inclined towards safer, low risk behaviour (Figure 11).

I would be happy for the self-driving bus to take more risks (e.g. pulling out into a smaller gap between traffic at a junction) to catch up time if it had been delayed in a traffic jam.

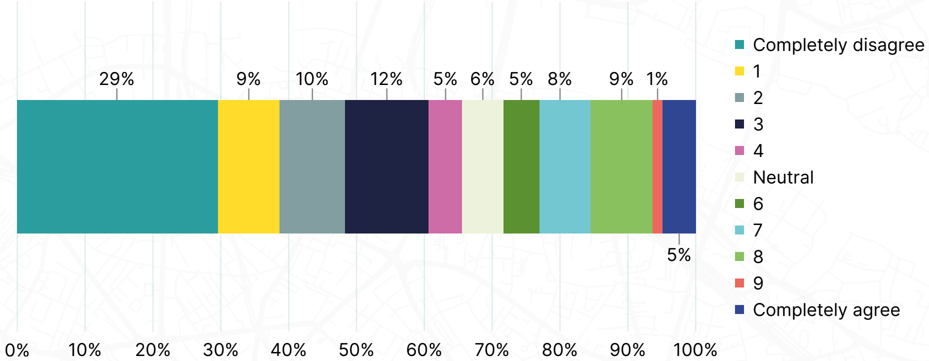

Figure 11. Risk taking to save time (mean = 3.39).

Safety (passengers vs other road users)

This theme explored perceptions of respondents towards self-driving bus risk management regarding safety, and the extent to which behaviours should be modified to protect passengers or other road users. Responses under this theme garnered little agreement across the sample, highlighting a diversity of views on the topic (Figure 12). The most frequent response (25%) was in the central position while 14.6% selected the two categories indicating the least agreement with the statement while 13.7% selected the two categories indicating the most agreement. There were no differences by sex or age.

When a similar scenario was presented as a dilemma between a pedestrian as a vulnerable road user, and the passenger of the self-driving bus, the survey highlighted a similar lack of agreement as to the type of behaviour that should be promoted (Figure 13).

If a collision with another vehicle is unavoidable, the self-driving bus should try to protect its passengers as its top priority, even at theexpense of other road users.

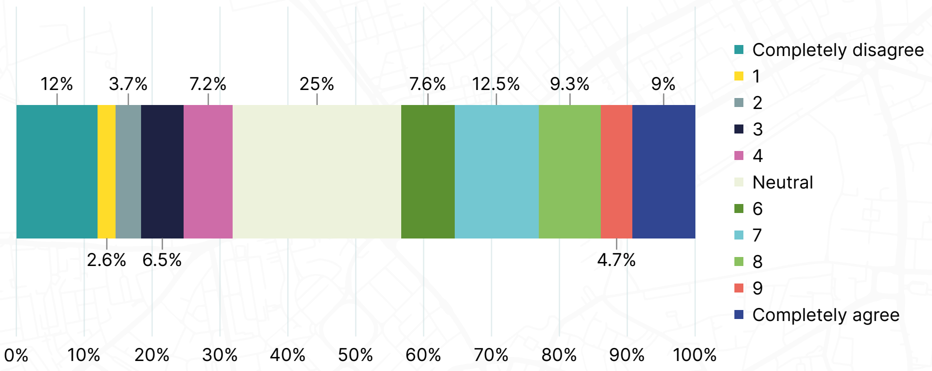

Figure 12. Safety response in the event of an accident (passenger vs road user) (mean = 5.23).

I would want the self-driving bus to take emergency action to avoid a pedestrian who unexpectedly stepped in the road, even if it meant risking injury to passengers on board the vehicle.

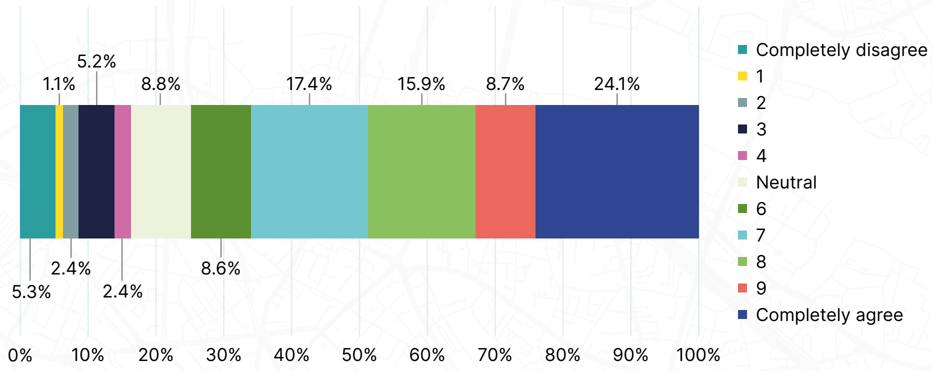

Figure 13. Emergency action response (passenger vs pedestrian) (mean = 6.96)

Rule breaking

Two statements specifically queried participants' views on whether the self-driving bus should break the rules of the road. The first asked whether the bus should not break the rules of the road to avoid holding up traffic – with the example of crossing double white lines in the road to overtake a fast-moving cyclist.

Figure 14 shows strong agreement that an SDV should not break the rules of the road to prevent holding up traffic. However, the next statement asked about rule compliance and avoiding a collision, asking whether the bus should break the rules of road to avoid a collision – with the example of crossing double white lines to avoid a pedestrian who had stepped in the road.

Figure 15 shows that participants were near equally convinced that the SDV should break the rules of the road to avoid a collision – highlighting that rule compliance is contextual.

The self-driving bus should not break the rules of the road to avoid holding up traffic (e.g. crossing double white lines to overtake a cyclist who is travelling at 14mph).

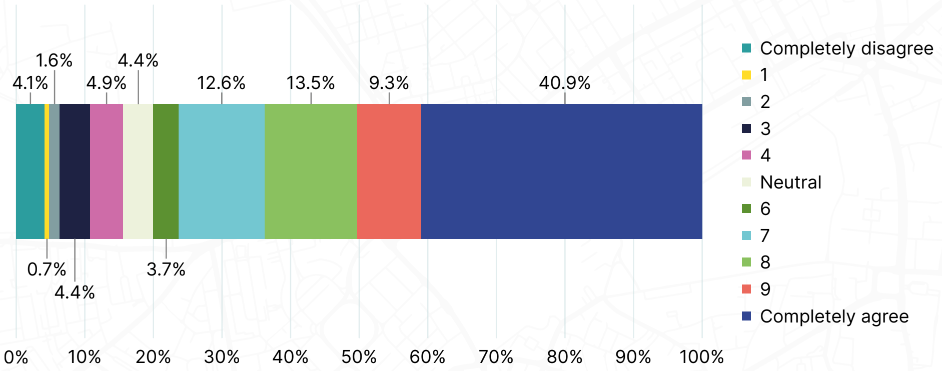

Figure 14. The self-driving bus should not break the rules of the road to avoid holding up traffic (mean = 7.68).

The self-driving bus should break the rules of the road if it means avoiding a collision (e.g. swerving across double white lines to avoid a pedestrian who has stepped into the road).

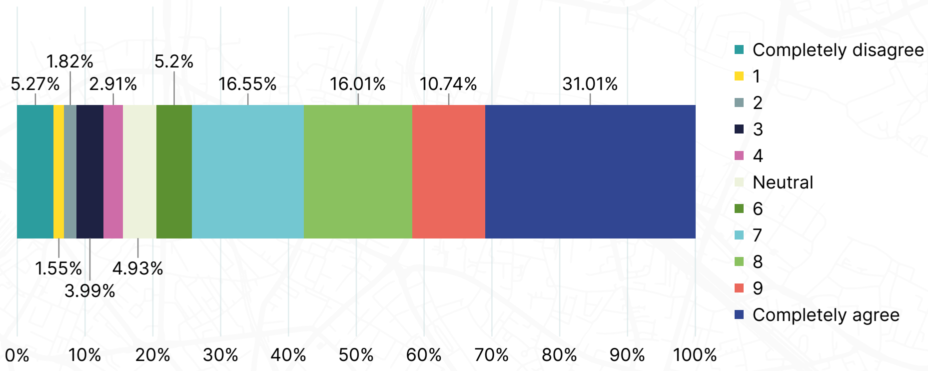

Figure 15. The self-driving bus should break the rules of the road if it means avoiding a collision (Mean = 7.32).

Prioritisation

A key requirement for defining an ethical goal function is to determine how an SDV should prioritise between the various risks and objectives that must be managed. The survey results were used to give an estimate of this prioritisation by applying positive or negative weightings to each statement with respect to one or more of eight ethical factors (see Appendix C). The results are shown in Figure 16, which shows the level of positivity towards each factor. This approach yields a range of -10 to +10 where +10 would be recorded if participants were fully in agreement with all statements assigned a positive weighting for each factor and -10 would be recorded if participants were fully in agreement with all statements assigned a negative weighting for each factor.

Figure 16. Estimated prioritisation of ethical factors from survey responses.

Figure 16 shows that 'Trust' was the most important factor from the survey results. This comprised statements relating SDV behaviours, actions by the vehicle manufacturers / technology developers, actions by the government to assure SDV safety and features necessary to support passenger safety and security. Interestingly, 'Safety of others' had a slightly higher response than 'Safety of passengers', indicating that survey respondents were at least as interested in the welfare of other road users than that of SDV users. 'Cost' was rated as the lowest priority but that is perhaps to be expected with participants indicating that developers should not cut costs at the expense of safety.

More surprising was that 'Legality' appeared to be low in the prioritisation. It appears that this reflects statements where participants were asked whether the SDV should break the rules of the road to avoid a collision or to allow an emergency vehicle to pass – where respondents tended to favour non-compliance in the interests of safety.

Workshop insights

Exercise 1: Risk-taking behaviour scenario

The following scenario was shared with participants, and formed the base for a series of follow on questions explore attitudes towards risk and safety dimensions:

"You are on a self-driving bus going to the theatre, but it has got stuck in a traffic jam behind several vehicles trying to emerge from a busy junction. The bus is programmed to take more risks to stay on time. Traffic is building up behind the self-driving bus. When it is ready to enter the junction to turn right, the bus slowly edges into the oncoming traffic."

Follow-on questions explored attitudes towards the scenario and looked to draw out differences in opinion as to the appropriate behaviour of the self-driving bus.

Findings:

Time versus safety: would the bus get stuck?

Participants noted in the scenario that there should be some expectation on the bus to move forward to emerge at the junction, otherwise it may not proceed on the journey at all if it were over cautious or unable to adopt an active role in the environment:

"I can see the safety concerns; the bus will be programmed to be on time, as a driver you could get stuck and be there for ever - so the right level of risk is therefore important." Workshop 1 participant.

Similarly, a participant in the second workshop noted that bus drivers are trained to adopt a driving style that suits the context given the types of situations/scenarios that occur in busy urban settings:

"I was on a bus yesterday and I saw a junction ahead that was blocked up. If the bus that i was on… if he hadn't edged out, I'd still be on the bus! The level of risk - if he moved out very slowly and I didn't feel a particular risk, I think - the traffic was built up - he wouldn't have moved. The bus must take some risks." Workshop 2 participant

Trust in the technology and manufacturers:

As part of Exercise 1: Risk taking behaviour respondents reflected that to be able to trust the vehicle in its behaviours and intentions, it was important to understand the training process for developing the AI algorithm – and that trust in the vehicle translates to trust in the individual or process training the algorithm:

"To be able to trust the tech behind the vehicle, the person who programs it will have made the right decision so therefore I can trust how it operates – because I can trust the vehicle." Workshop 1 participant

Trust in the bus operator was also raised. In the event of the vehicle having to emerge into the busy road participants noted that they would have to trust the bus operator to be able to use the service, and that in the event described it is important that the vehicle operates within the rules of the road:

"Bus must stick to the rules of the road – can't break them." Workshop 1 participant

Trust in the manufacturer of the self-driving bus was also raised by participants, with differing views being shared as to whether manufacturers could be trusted. Whilst some thought they could be trusted:

"I think (the bus) should stick to the rules – I would trust the manufacturer would operate in every situation." Workshop 1 participant.

Others in the workshop were more concerned given their own experiences of other new technologies suggesting that incident causation can sometimes be unfairly attributed to the driver when the system may have been responsible for the error:

"I don't trust the manufacturers – some tech has not been well developed (e.g. assisted driving technology) – I think vehicle manufacturers are always placing fault on the driver - even though the technology is assisting them, and sometimes malfunctions. I don't think that is right." Workshop 1 participant.

Trust in the company running the service was discussed between participants in the second workshop who considered whether or not the service operator should have extensive control over the behaviour of the vehicle, or whether that instead should be determined by another party, e.g. a central body. This they thought would mean there could be more trust in the service operator:

"I would question this – would it be up to the bus company? Shouldn't the behaviour be (determined) from some input from a governing body?" Workshop 2 participant.

RESULTS

A recurring reflection from participants was the low level of trust in the automobile industry following the VW emissions scandal (see Jung & Sharon, 2019). Those at workshops noted that any new services would need to exist under a strong governance framework, and that there would need to be a strong framework and regulations. This also highlighted for one participant the challenge when governments can also no longer be trusted:

"For example, we can't trust big car companies, like those that cheated environment figures (VW) – none of us expected that but now we know it happens. Emissions scandal shows that the companies are pushing the boundaries and not following the law." Workshop 2 participant.

"But what if the law is wrong? If you have a corrupt government, you can't trust them to make the law?" Workshop 2 participant.

Safety was paramount.

Safety emerged from this scenario as the most important factor to be considered, over other issues such as time and cost. The self-driving bus was deemed to have to avoid making overly risky manoeuvres to be able to meet its objective of arriving on time, and should prioritise both the safety of passengers and the safety of other road users:

"It's important to save lives over time – I see why people would want to be on time. Time isn't most important; you shouldn't risk other people's safety. Shouldn't risk other people's safety who are on the road either."
Workshop 1 participant

Another red line that emerged from the discussion was the need for the new self-driving bus service to have a safety standard at least comparable or better than existing human driven services. There was no appetite for a vehicle that was less safe than existing vehicles, but has improved qualities, such as low cost, or low emissions. In this case an improved safety standard was a clear ethical red line.

"Greatest risk is not caused by the self-driving vehicle, but by the other person driving a non-autonomous vehicle. If it's two self-driving vehicles, then they will communicate with each other and there will be no collision. But if it's a person who is stressed (e.g., fighting with another person in the car) then you don't know what could happen. No way would I risk my life just to get there on time."
Workshop 2 participant

Risk appetite is individual.

Defining an appropriate level of risk was difficult for some participants who recognised differences in individual attitudes to risk level – some were comfortable with a more active driving style whilst others preferred to be more cautious in the scenario described:

"Risk is not absolute - I would take more risk as I am comfortable with the risk. (I think) where it says "slowly moving into the road"; that is a key fact - e.g. the slow behaviour is a smaller risk. I would be concerned if the bus were not moving slowly into the road."
Workshop 1 participant

Another participant noted that there would need to be a level of risk analysis by the vehicle that would mean that risks are weighed and understood before decisions are made, like the cognitive process a human driver would go through:

"I naively assume that even if the bus is programmed to take more risks, I assume that the bus will have some kind of protective mechanism to take small risks, not big risks. It would have to assess the size of the risk."
Workshop 2 participant

The concept of risk of harm was key to one participant, who highlighted that there are other types of risk associated with the vehicle's operation, but it is risk of harm to individuals that should be the major risk to be managed and mitigated against:

"In my view risk is the key word - edging out in traffic it should be risk. There is a risk of some harm. Its whether there should be a risk of harm from the car emerging. What "risk of harm" is to someone is key in this."
Workshop 2 participant

RESULTS

Legal behaviour overrides other behaviours, even if they are potentially going to save time.

One respondent recalled an incident in which a human driven bus was driven illegally and created a higher risk to other road users than the respondent was happy with. In this instance it was deemed important that the self-driving bus behave legally, even if it means the bus is late against its schedule:

"I have an experience of a bus acting illegally: I was on a bus where the driver was stuck in a queue of traffic to turn left, and he pulled into an opposing lane to then take a left turn - it was an illegal manoeuvre; I wouldn't expect the automated vehicle to do that.
Road law means that you cannot go into the opposing road to go around the corner - as a road user, you should sit and wait. I would expect that the self-driving bus would sit and wait before moving forward and would not act illegally." Workshop 1 participant

Rule keeping was important to participants in Workshop 1 but participants were not convinced that self-driving buses could be trusted to follow the rules without any interaction with a human. Safety operators were important – described as a red line by one participant:

"I don't think I could trust a vehicle to operate fully alone, there needs to be someone involved in case something goes wrong." Workshop 1 participant

The legal dimension was further explored when an ambulance was included in the scenario for participants in workshop 2. In this instance participants were asked to consider whether the vehicle should be able to make an illegal road manoeuvre to make way for the emergency vehicle – in this case, moving into a crossing on a red-light, to allow an ambulance to pass. In most instances participants agreed that legal manoeuvres were most important:

"I don't think the bus should accept more risk just to let an ambulance through." Workshop 2 participant

"Is it legal on the road today? I think at present you are unlikely to be prosecuted but you will commit an offence to get out of the way, e.g., at a red light." Workshop 2 participant

"It depends on the risk of harm for those in the ambulance versus the riders on the bus." Workshop 2 participant

One participant noted that the legal framework for the behaviour of the self-driving bus should be set, and the vehicle be adapted to fit its requirements. There was concern that the implementation of new self-driving services could shape existing laws, which one participant saw as concerning:

"(For) other forms of transport laws can be changed - transport is a dynamic environment...government policy should be shaped by vehicles, and existing road infrastructure: one self-driving bus should not override all road law." Workshop 1 participant

Other road users:

Participants also noted that other road users were the source of risk in this scenario particularly as the vehicle was looking to emerge, other drivers, cyclists or pedestrian may not be confident around a self-driving bus, and could therefore act in an unpredictable manner:

"Seeing as all the other road users are human then I think the risk is appropriate - Everyone else is also highly risky." Workshop 1 participant.

"My issue with AVs is the human error side of things (but) by the time they are commonplace then the error will be eliminated to the point where they are safer. You can't rule out pedestrians and other vehicles and tech malfunctions." Workshop 1 participant.

Standards

The behaviour and risk appetite of different operators was a subject of discussion for the group: one respondent noted that they expected that all operators would need to adhere to a set of standards that would over time mean that all operators can respond to risks as and when they arise:

"Won't different operators have different risk levels? Will you get a feel for the type of risks they are willing to take? Or will they be standardized? As an operating company you should be able to tailor to different risky situations. If the operator couldn't change the risk levels, they can accept then I would trust them less." Workshop 1 participant

Participants also noted the tension between standardising behaviour for all contexts or adapting the behaviour to meet local norms. The example of London driving style versus other parts of the UK came up, where it is commonly believe that a more active/aggressive driving style needs to be adopted by drivers:

"You will need a different design in London, where you need to drive aggressively. I would expect in London a bus driver will take more risks, but in the countryside, you wouldn't want the same level of risk behaviour from a bus." Workshop 1 participant

Accountability

Participants in workshop 1 were concerned about the accountability of the service should an incident occur. The lack of a clear party to blame meant that there was some concern over how decisions were to be made, and if they were the right decision:

"The issue I have is that I don't know who to blame for the decision that the self-driving bus makes – a bus shouldn't compromise other people who are involved (e.g., other road users). All road users will have some involvement – if you are basing decisions on an algorithm - aren't you trading lives?" Workshop 1 participant.

One participant believed that the behaviour of the vehicle should be like that of a human driver in the same situation, and it should drive according to the standards of the context – in this case the participant believed the vehicle should adopt a more active driving style:

"Destination is a red herring: what we're talking about is safety. The bus isn't acting any differently to any other person who drives in London. I would approach it very carefully and edge out until someone stops. I'd want it to do the same as I would do as a driver." Workshop 2 participant.

Driving as a learning process

There was acceptance among participants in Workshop 2 of the need for the self-driving bus to learn through experience, and that there will need to be some acceptance of mistakes by communities at the beginning of its use. Participants agreed that this should be allowed, so long as vehicles learn over time. It was not discussed however whether the training of vehicles should be shared between manufacturers for the greater good of training services in general.

"Mistakes are a teaching process to get it right." Workshop 2 participant.

"I do think that when the bus is first introduced there will be some mistakes; this is what happens when something new is introduced. But they will learn over time." Workshop 2 participant.

Ethical red line: in this exercise both workshop groups were uncomfortable with the self-driving bus making risky decisions to save time and keep on schedule. It is unclear how participants would view this behaviour in a different scenario (e.g. a work appointment rather than a leisure appointment), but it is clear that within this particular scenario safety was a top priority, above others including timeliness.

Ethical red line: this exercise also highlighted that there is an emerging red line related to operating within a clear legal framework, and the importance of adhering to the rules of the road under all circumstances. A high level legal and governance framework was deemed important by participants, not only for the purposes of safe operations, but also to ensure that users are able to trust service providers and vehicle manufacturers:

"My red line is that there must be a clear legal framework. We need a legal framework that is set by society and not by the manufacturers of the vehicle. I wouldn't trust it if it was from the manufacturers." Workshop 2 participant.

"Should not break the law – but I can think of scenarios where I want them to break the law. Say a traffic light breaks and is stuck on red, then the vehicle will never move - but that is not good. What would it do if other issues occur on the road?" Workshop 2 participant.

Exercise 2: Road safety scenario

The following scenario was shared with participants, and formed the base for a series of follow on questions explore attitudes towards legal and safety of others dimensions:

"You are standing on a crowded self-driving bus as it moves at 30 mph through the city. The bus needs to brake suddenly to avoid causing serious injury to an elderly person crossing the road - this is a shock to you and the other passengers as you experience a sharp jolt and lose your balance."

Findings:

Safety of the individual vs safety of others

Participants were divided over whether there was a need for the vehicle to adapt its behaviours according to the perceived vulnerability of different road users, e.g., according to age or mobility requirements. For some participants this was considered inappropriate and potentially unfair:

"Weighting the severity of (the decision) e.g. other people may have less severe injury such as a younger person – I don't think that should change how the vehicle behaves." Workshop 1 participant.

Whilst others recognised that "weighing the lives" of others is something bus riders use do at present when they assess the risk of using a service, and therefore why should the self-driving bus behave differently:

"Do we want people to weigh lives equally, even if (they are) 20 years old, compared to an elderly person? You assume the risk when you get on to the bus now. If someone gets on now, you assess the risk already. I think everyone is assessing the risk all the time." Workshop 1 participant

There was also division as to whether the safety of passengers should be prioritised over the safety of other road users, and whether the vulnerability of different service and road users should be considered by the self-driving bus. Whilst one participant considered the safety of other road users a priority, another believed that the self-driving bus should value all road users equally:

"More important to protect the person on the road as they are more likely to be injured." Workshop 1 participant

"I would accept the injury on the bus, to mean that the other road user is protected." Workshop 2 participant

Accepting injury to self-driving bus passengers over passengers has implications for the internal design of the vehicle. OEMs will need to ensure that the interior of the vehicle is suitable to this scenario. This was highlighted by participants:

"I do think it's the responsibility of the bus operator to make sure the internal bus environment is safe (e.g., has a lot of handrails)." Workshop 2 participant

"As a bus rider – I have a responsibility to look after my own responsibility - I should do that as much as I can – but the bus operator must make sure I can do that myself too." Workshop 2 participant

When the pedestrian in the scenario was changed to a dog, there was disagreement between participants to how best the vehicle should behave:

"I would benefit the dog – I would want to stop for the dog, even if it meant people on the bus were injured." Workshop 1 participant

"Save people on the bus over the dog on the road." Workshop 1 participant

"I would choose the dog – it would be stressful for me to be on the bus to see, I would want it to brake heavily, I wouldn't like the fact the dog was hurt." Workshop 1 participant

"Would expect the self-driving bus to behave as a driver would: i.e. if it puts people in danger – if that dog runs out and you're travelling at 30 mph and you brake, you are more likely to injure the people on the road. A human driver will take reasonable measures to not hit the dog." Workshop 1 participant

"I would want the bus to prioritise the passenger – you have to have the safest approach." Workshop 1 participant

This difference was further made apparent when an additional actor was added to the scenario, in this instance a close following moped that, should the vehicle stop suddenly to avoid the dog on the road, could collide with the back of the self-driving bus. In this scenario there were clear values-based decisions that came to the fore for some participants, who viewed the illegal behaviour of the moped driver as reason enough to place them at greater risk than the dog on the street:

"I would prefer to put the moped at risk over the dog, because the moped is not behaving legally. It's following too closely behind!" Workshop 1 participant.

Another participant noted a similar concern, highlighting that the rules of the road mean that those who operate outside are "fair game" to an incident:

"The most important dimension is safety. Individuals who are not behaving legally are fair game for an accident for that reason – therefore the bus should run over the dog over protecting the illegally behaving moped." Workshop 1 participant.

Risk

Risks associated with other road users were highlighted in the workshops, particularly during discussions regarding moving in busy urban environments. Other drivers and pedestrians and were cited as a major source of risk to the self-driving bus, and those using it:

"With a driverless bus the risk of failure is less (but) the human error of other road users plays a key part. Slamming on the breaks when someone steps in the road, someone (on the bus) may fall over – that isn't good. Risk to other road users: that is less likely to happen in my opinion." Workshop 1 participant

A common theme throughout the discussions was the general perception that current driving standards are low, with participants noting that they regularly experience bad driving, or unsafe road use by cyclists and pedestrians. A regular source of risk to participants were other road users, and not necessarily the self-driving bus which some in attendance trusted to make appropriate decisions.

One participant noted that the self-driving bus should behave in a highly risk-averse manner, with the expectation that it would drive slow enough to avoid any incident that may occur: driving slower than current vehicles and ensuring the safety of passengers on the vehicle:

"In my view the self-driving bus should always travel safely, sometimes as slowly as possible – to make sure it's as safe for passengers as possible." Workshop 1 participant

Costs vs benefits

Costs were one of the clearest lines to emerge from the workshops. This resulted in discussions around people's ability to pay for a safer service and the idea that people of lower incomes should not have to make a choice between mobility access and safety. When asked whether lower safety standards were acceptable for lower costs, all participants responded negatively, highlighting the need for bare minimum standards to which all services should adhere:

"I would not like cheaper rides if it meant it was going to be a risky bus. Safety is my main concern over the price of the bus." Workshop 1 participant

"I wouldn't accept lower safety standards even if the self-driving bus was cheaper." Workshop 1 participant

Trust in the vehicle

As part of this scenario participants explored the difference between trusting a human bus driver, and trusting the self-driving bus, when presented with the collision situation. On participant noted that the technology in the self-driving bus could be considered like London's Docklands Light Railway (DLR), a partially autonomous metro system:

"I am happy to ride on the DLR, and there is no driver on that." Workshop 2 participant

"It should be equal people on the bus and the people off the bus – everyone must be protected equally. It should not have to make the decision. Everyone should be treated the same." Workshop 2 participant

Ethical red line: Safety standards should not be influenced by the 'cost at the point of use' for the customer across different service offers. Participants in workshops were clear of their expectations for a broad high level of safety, that is greater than current standards, and which all service users have access too. There was clear agreement that at no point should a low-cost and reduced safety service be allowed to enter the market.

Exercise 3: Data privacy/data sharing scenario

The following scenario was shared with participants, and formed the base for a series of follow on questions explore attitudes towards trust and legal dimensions:

"A self-driving bus is travelling on a one-way street and crashes into a parked car to avoid hitting a pedestrian who has walked out into the road. None of the passengers are harmed and only the vehicles are damaged. The bus's on-board computer has recorded all the data that shows why the bus made this decision, the view from the on-board camera of the pedestrian stepping into the road and the passengers inside the bus."

Findings:

Data sharing

Participants were asked to consider whether or not data sharing between service operators, government and other institutional actors was appropriate in the event of an incident. The purpose of data collection and data protection was important to participants; some were aware of the GDPR and the obligations of those subject to it, and as such had expectations on self-driving bus services to operate in an appropriate manner.

Data sharing between service providers, technology developers and government was considered to have value, not only in understanding the specific of the incident (e.g. for insurance purposes) but also to support service and technology improvement for example increasing service safety, service reliability or user comfort:

"Data should only be kept for safety and costs, and for overall safety improvements but then be deleted afterwards." Workshop 1 participant

"(Data should) be shared with the authorities, any buses. would be happy to share it to prevent the same accident from happening again, as long as there's nothing personal." Workshop 2 participant

"All data should be used by the manufacturer and insurance company but important to be deleted if not anonymous." Workshop 1 participant

"Data must be used for a certain purpose only, improving how a bus is automated. Important bus company uses data to improve experience and minimise damage to people and others." Workshop 1 participant

One participant recognised a difference between vehicle manufacturer and operator, noting that different institutions should have different access to data according to their needs:

"I don't think the data should go to the manufacturer of the bus, but the operators should be able to use it." Workshop 1 participant

Regulation and governance

Data sharing was also considered to interact with perceptions of trust in the service. One respondent noted that although consumers may not wish to see all data, trust in service operators more generally would be improved by regular data sharing between service providers, manufacturers and regulators/government:

"In order for people to trust the tech the vehicle, its manufacturers would have to share info with the government." Workshop 1 participant

"Data should be passed on to the relevant authorities including insurance company." Workshop 2 participant

Some participants argued for the creation of an independent body from industry that investigated incidents and supported improvements across developers and service providers. Whilst some participants saw this as excessive, others viewed it as important for building wider trust in the system:

"(I think it's) important to have independent body to investigate any incident/collision. All data should go to independent body and be restricted to body unless necessary to involve the police and/or insurers. Even incidents with no personal harm this would be important, as operators and manufacturers have a vested interest." Workshop 2 participant

Other participants considered government inclusion in the process as excessive, and instead preferred for data sharing to occur between the service operator and the other party/ies involved in the incident:

"(I think) The manufacturer should just deal with it themselves. If you have a car crash you just swap numbers you don't get the government involved, so why should the government get involved at this stage?" Workshop 1 participant

"I'm open really. Not sure whether data should be shared with authorities." Workshop 2 participant

RESULTS

Accountability and individual rights

Participants noted that users/riders are unlikely to have much choice in how their data is used, other than the decision whether to use the service. One participant noted concerns over the extent to which the service provider is likely to require the public to agree to data policies, and the risks associated with them:

"Presumably when you buy a ticket you have to accept the terms and conditions - so you accept the way the company wants to use the data." Workshop 1 participant

"We've got no choice, for safety you need to pass the data on." Workshop 1 participant

Individuals involved in any recorded incidents were also considered to be potential recipients of data. This was considered by one participant to be an important right for individuals using the service, who may be able to hold the service operator or other actors accountable:

"If bus has recordings etc, they could be held accountable. Therefore, would be happy to share information for that reason." Workshop 1 participant

"People in accidents have a right to understand what happened in the event of a crash." Workshop 1 participant

Data sharing for service improvement

Participants in workshops noted that under certain circumstances it should be ensured that data is shared between developers and service providers to ensure that safety standards across the wider ecosystem improved. Participants recognised this as a key benefit of the connected self-driving services, and one which could lead to net improvements in safety.

"It should be mandated to share all of your data, to be an operator to enable to greater good." Workshop 2 participant

"Land Rover changed the front of the Land Rover Discovery because it was found to be too dangerous. They looked at injury data which was recorded, the car model data which was recorded and decided to change the design. Data is important, and it really needs to go to everyone involved for the greater good." Workshop 2 participant

Ethical red line: Privacy and data sharing – participants in the workshops were comfortable with their data being used to improve safety and improve services and saw this as standard for the use of modern public transport. Given our sample of urban dwellers in London, who frequently use public transport, this is perhaps unsurprising, and points to a general acceptance of data processing and data sharing for the use of public transport services. The workshops did however highlight a potential red line regarding data privacy for participants who viewed their own personal data (e.g., images of faces) as information that should not be processed or shared between services. Not all participants felt strongly about this point, but there was no significant disagreement that services should protect individual identity.

Participants considered data sharing reasonable should an individual be harmed in an incident. In this case we can define a red line as to data sharing between reasonable parties, for example insurance companies and vehicle manufacturers.

Participants felt it important that there is transparency as to who receives personal information in these instances, and that there is clarity as to the nature of the processing of data. Insurers, police, and legal entities were all expected to make use of this data.

It was clear however that participants were not comfortable with personal data being sold or used for marketing purposes. This was a red line for many participants who noted that advertising is pervasive and intrusive – but also accepted in some contexts.

Ethical red line: Data sharing is allowable, only to improve safety generally – data can be shared to understand incidents and to improve safety, but operators must do so in line with the laws around data privacy and data protection.

Road safety benefits discussion

Part of the workshop explored perceptions of the potential safety benefits of the self-driving bus, introducing the idea that the self-driving bus and other self-driving technologies is likely to lead to a net decrease in collisions and road fatalities.

Participants were asked to consider this information, and then to participate in a facilitated discussion to understand public sentiment towards road safety, and the potential role of the self-driving bus. Probing questions explored (1) emotional reactions to the potential safety benefits (2) acceptance of self-driving buses when there is no perceived safety benefit; and, (3) acceptance should safety statistics worsen (e.g. number of incidents and/or fatalities increases)

Findings

Safety improvements are considered a major benefit by participants: Participants felt it was important to recognise a reduction in incidents as a benefit of the self-driving bus.

“It's a benefit that there will be less fatalities – that has to be a good thing.” Workshop 2 participant

There was a clear expectation that there should be a net reduction in road fatalities, for example. However, it was unclear whether there is a target level to which the self-driving bus should look to achieve:

"There are always going to be fatalities on road. I would expect there to be fewer. Policy makers will expect there to be some – that is just a fact of life. For an individual though, every death is important. It could be someone you love” Workshop 1 participant

"There will always be fatalities, you can’t get away from that. I think with self-driving buses it will be pedestrian error rather than vehicle error. I would expect that self-driving buses have the highest safety standards" Workshop 1 participant

"(The self-driving bus) will never be 100% safe. Plane crashes are very unlikely for example. (I think that) overall self-driving buses could reduce fatalities by 90% for example - it’s an overall decrease – e.g., you are valuing everyone equally." Workshop 1 participant

There was also an expectation that there must be at least some net positive safety benefits of the self-driving bus. It was unacceptable to all respondents that there could be an increase in road incidents because of self-driving buses:

“I would not accept (lower safety standards), I would not use them. Every box must be ticked for me to get on one, or let my daughter get on them.” Workshop 1 participant

There were perceived risks of safety standards of self-driving bus services, including the potential for a two-tier system to develop for public transport:

“If it was my only option, I would get on it – but my choice would be to not take the jobs away from people. In terms of lower safety standards – I would not want a 2-tier system, you would value people’s lives less for those who use this service. It could be a two-tier system because people who could afford one service would get a better service, and those who couldn’t afford it may get a lower cost and unsafe service” Workshop 1 participant

RESULTS

Safety of vulnerable road users was also a clear issue for some, who when reflecting on the safety statistics felt that there is a difference between different road users and their vulnerabilities:

"I do think fatalities are different: there's killing children and there is killing adults. Therefore, the word fatalities need to be clearer." Workshop 2 participant.

"What if they were not very good at predicting child behaviour on the road, e.g., if it was crossing the road? What if more child deaths went up? I don't think I would accept that." Workshop 2 participant.

Ethical red line: Safety – Discussion around the safety statistics highlighted that there are several red lines relating to personal safety and safety of others that are important to the public. Firstly, there is a contextual red line, in that there is an incidence level that is important to the public that determines acceptability. However, when probed participants couldn't quantify a specific number or proportion of KSIs.

The second red line is that the self-driving bus service should protect all road users, no matter what their needs or values. This wasn't important to all participants, who viewed illegal road behaviour as resulting in individuals becoming 'fair game' to an incident. However, others considered that all road users should be protected:

"It must protect everyone, even stupid people who are crossing the road. Everyone should be cared for." Workshop 2 participant

A consistently held view by participants was that the vehicle would need to demonstrate similar or a net improvement in road safety statistics – road safety cannot reduce because of the introduction of the vehicles:

"I would accept it if it was no worse – I would get on it." Workshop 2 participant

"Surely there will be people who will be using the data to improve the service – I would want there to be improvements in the system, if it isn't learning, then what is the point?" Workshop 2 participant

"No – I'd rather pay a few quid more than accept worse safety standards than what we have now." Workshop 2 participant

"If it were the only option I would find another way – I wouldn't use it if it wasn't as safe as what we have today." Workshop 2 participant

"It is making the journey anyway and it's not going to make a difference to other users then I would still use it. But if the safety standard went down and affected the other road users – if it were affecting general road users, and I feared for my own safety, then I wouldn't get on it. It doesn't make a difference if I ride it though." Workshop 2 participant

RESULTS

Exercise 4: Dimension ranking exercise.

This exercise explored the potential for ranking the importance of different dimensions explored throughout the session. Participants were provided with a set of dimensions to order from 1 (most important) to 8 (least important) when considering the development of a new self-driving bus service. The ranking results for each participant are shown in Figure 17.

	Workshop 1								Workshop 2						
Participant:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**
Fairness: self-driving vehicles operating in a manner that treats road users fairly. For example; giving more space to pedestrians than other vehicles	6	5	4	6	4	5	5	4	6	1	4	7	5	7	6
Mobility: the ability for you and people in your community to travel easily to work, school, hospital to see friends etc	4	6	8	5	5	7	6	5	5	6	6	6	1	6	7
Legality: self-driving vehicles must obey the rules of the road at all times	1	2	2	1	1	4	4	8	2	4	5	3	6	2	4
Safety of passengers: operating a self-driving vehicle to protect the safety of passengers within the self-driving vehicle	3	3	3	2	3	2	1	2	7	3	2	5	4	5	5
Trust: trust that self-driving vehicle manufacturers and government regulators have done everything necessary to ensure that self-driving vehicles are safe and ready for the road	2	1	7	4	6	1	2	3	3	5	3	1	2	3	2
Safety of others: operating a self-driving vehicle to protect the safety of other road users such as pedestrians and cyclists	5	3	5	7	2	3	3	1	1	1	1	2	3	4	3
Cost: the full cost of delivering the self-driving vehicle service, including costs of development, costs of keeping safety data records, inusrance costs, value of intellectual property. Higher costs of development and operation may means higher ticket prices and/or less frequent services	7	7	6	8	7	6	8	7	4	7	7	4	8	8	8
Urban design: the extent to which we need to adapt our streets to accommodate self-driving vehicles such as having their own lanes and signage, barriers to prevent pedestrians crossing in front of them etc.	8	8	1	3	8	8	7	6	8	8	8	8	7	1	1

Figure 17. Dimension rankings

There was tension between which different ethical dimensions participants felt were most important in the workshops:

Safety of others
(Mean ranking: 2.93; highest ranking: 1; lowest ranking: 7)

This dimension was regularly given high rankings by participants in the workshops, who noted the importance of protecting vulnerable road users through the adoption of self-driving bus services:

"I put the safety of others as most important, because it is the starting point (of a good service) but I found it really hard to differentiate, they all are important in different ways." Workshop 2 participant

Trust
(Mean ranking: 3.00; highest ranking: 1; lowest ranking: 7)

Participants rated trust highly when ranking dimensions, citing its importance as a state for encouraging the use of services. Participants felt that without trust it is unlikely that the public will use new services as and when they become available:

"Unless this is the case (e.g. there is trust), it's never going to get off the ground. It was quite difficult to choose, there are a lot of contenders, safety for example you could argue that they should be first and equally cost too – you could argue without that it's not going to happen." Workshop 2 participant

Legality
(Mean ranking: 3.27; highest ranking: 1; lowest ranking: 8)

Participants strongly asserted that the self-driving bus should stick to the rules of the road, resulting in its emerging as the third highest ranked dimension across the two workshops:

"It's very important that the bus sticks to the rules of the road." Workshop 1 participant.

"The bus must stick to the rules of the road – it can't break them" Workshop 1 participant.

"The rules of the road are sacred almost because they ensure safety" Workshop 2 participant.

However, this contrasts with the survey results in which participants strongly suggested that an SDV should break the rules of the road to avoid a collision with a pedestrian (see Figure 15).

Safety of passengers
(Mean ranking: 3.33; highest ranking: 1; lowest ranking: 7)

Only one participant put the safety of passengers as their highest ranked category but most placed it in the top-half of their ranking of all dimensions.

"In my view the self-driving bus should always travel safely, sometimes as slowly as possible - to make sure it's as safe for passengers as possible" Workshop 1 participant.

However, it was noted that the safety of others should be seen as more important since other road users had not made the positive decision to board the bus but were exposed to risk from its operation:

"I put the safety of others first because they have not made a choice to use the bus so they need to be considered more because they haven't chosen to participate in the risk" Workshop 1 participant

Fairness
(Mean ranking: 5.00; highest ranking: 1; lowest ranking: 7)

The complexity of fairness versus legality dimensions meant that they were difficult for participants to apply – this meant that one participant instead prioritised safety above the other dimensions on offer:

"Fairness is hard because it is subjective. Legality is hard because its fixed and it's not compatible with the dynamics of how road situations change, therefore I decided to prioritise the safety of passengers" Workshop 1 participant.

One participant reflected that there is now some level of fairness built into the regulations in the form of priorities on the road:

"Some fairness is already built into the legal system now with the road hierarchies." Workshop 1 participant

One participant noted the relationship between fairness and trust in arriving at their rankings:

"They go hand in hand, the fairness and trust dimensions feel closely linked" Workshop 1 participant

Mobility
(Mean ranking: 5.53; highest ranking: 1; lowest ranking: 8)

Mobility was ranked moderate to less important compared to other dimensions, with participants highlighting it as a useful dimension, but not as important as others (e.g., safety of others). Mobility was considered to overlap with fairness, and participants highlighted perceived similarity between the two:

"I think mobility should be high. Bus operators need to know who the community is, and what they need from the self-driving bus." Workshop 2 participant

Urban design
(Mean ranking: 6.00; highest ranking: 1; lowest ranking: 8)

Urban design drew the greatest range of responses across participants (voted most or least important by participants in both workshops). When cited as the most important dimension, it was highlighted that the adoption of self-driving buses is unlikely to work if there is not adequate consideration to the design of urban spaces:

"None of this is going to work unless the environment is set up for self-driving. It was clear that this was the most important consideration for me. The environment has to be conducive." Workshop 2 participant

"I put urban design on the bottom. I question if we need to have special dedicated lanes, then in lots of London streets and roads there won't be space for other vehicles. Therefore is it feasible in our current road system?" Workshop 2 participant

Cost
(Mean ranking: 6.8; highest ranking: 4; lowest ranking: 8)

Cost was cited as least important in driving decision making, with some participants placing it last in their preference lists. Reasons appeared to centre on the need to ensure safety above all else. There was consensus among participants that safer services should not be more expensive (e.g. all should have access to the same level of safety standards, no matter their purchasing power).

There were however concerns over the right market rate for using the service, and some seeing cost as a useful barrier to entry:

"What is the range of costs associated with the self-driving bus service? The price at which something is placed changes other decisions i.e. if it's free there might be too many people using it." Workshop 1 participant

Others worried that a free service to all users might mean that important user data that can support ride safety, may be lost as riders may not need to log their journeys:

"If something is free, then people don't have to tap in and there's no log of passengers if something goes wrong." Workshop 1 participant

Others believe that cost was not the most important factor in driving self-driving bus adoption. There were some expectations that cost would be high initially, but some concluded that the cost of services should decrease over time as the technology becomes more mainstream.

"Things are going to get cheaper for AVs, initially it will be expensive. and once it starts running it will get cheap. Cost is therefore not as important." Workshop 2 participant.

Sentiment thermometer: reflection exercise

The first and last exercise of the session was designed for participants to reflect on their experience of the session and to draw out any ideas or views that they felt had changed over the course of the discussion. Using post-it notes participants noted where they positioned themselves on a scale of 1-10, where 1 = Fully unconvinced that self-driving services will benefit me and my community; and where 10 = Fully convinced that self-driving services will benefit me and my community. Individuals made notes separately and shared their views, including reasoning for their decision:

RESULTS

Table 1. Workshop1 - Sentiment thermometer findings.

	Start		End
"I think self-driving can benefit the community and I think it's possible."	8	"The workshop has raised a lot of questions in terms of impact for other types of road users that are not me- how decisions around safety and fairness for others and the decision process for reporting issues."	7
"I'm middle of the road, and an ex police officer I've experienced lots of challenging incidents on buses...who do you report it to if the bus is autonomous?"	5	"We've got to embrace change, and progress is important- if everything is done properly. Reservations will not change; I am still placing trust."	8
"As long as there is a driver who can take charge in case of any failures, I'm convinced its workable."	6	"Transport of the future. I hope we'll get there. To move to an 8, convince me that there will be an operator on the bus."	7
"I expect them to be safer than regular vehicles concerned about them taking jobs, but there are lots of bad drivers out there"	7	"It could be used to benefit rural communities and if data could be used to monitor and make the best bus routes (could be more dynamic), accessibility for others is important. I believe in terms of safety it'll be safer."	7
"There's enough idiots on the road without having people not realise there are driverless cars"	1	"I don't think a self-driving bus could benefit me at all. I'd prefer council spend money on improving other transport my area."	1
"On the basis that it works as it will greatly benefit many people, mobility issues, people won't be late."	9	"If there is a moped behind the bus, how is the bus going to think- what is the right decision in that situation how is it possible to compromise in these situations? Unless we engineer the bus to think in the way we think... if we have to put another lane in for AVs, we can barely squeeze our cars through on the road anyway, we can't make the road any bigger. I don't accept (change to) urban design. If they can all share the same road that would be better."	8
"As a teenager I think they will be a big part of my life in the future but I'm still sceptical"	4	"I've been able to see a lot more how regulation is really a struggle and how something universal is unachievable at present. The idea of compromise is an issue – its difficult to quantify the ethical needs and considerations as a human can do. AI (capability) is limited. Also, how useful is it to learn to drive if AVs will come through?"	2
"Severe reservations about the tech but conceptually can see that it will benefit the community, they look better than current buses."	7	"It's a great concept but I Don't think execution will be as it should be."	7
Mean = 5.88; Range = 8		**Mean = 5.88 ; Range = 7**	

Table 2. Workshop 2 – Sentiment thermometer findings

	Start		End
"Excited for the future and considering the progress that has been made so far I have every confidence that even though people are scared now, it's going to be a wonderful development."	10	"Get on with it!"	10
"It's scary commuting every day in London. Self-driving feels like it could be safer. they've got to be better for the environment. might bring more jobs and be good for the economy	7	"We all think about accidents first, but the urban design and mobility are important."	7
"I wouldn't' be at today's session if I didn't think they're a good idea. I do see issues about the community and environment where self-driving vehicles exist. I also see some moral challenges in their adoption, but I'm cautiously optimistic."	7	"I still feel as confident. maybe slightly more but not enough to go up to 8. Had not thought about all the different issues which thinking about may affect how I feel about it and how I think it should be introduced."	7
"I'm unsure – I want to learn more this morning. keen to see how he feels later in the day"	5	"I don't think my community would benefit much from the bus service mainly because the roads and people - London is not set up for it. in a different community and if you were starting from scratch and building a new town then fine, but to adapt this tech for SE London is a step too far."	6
"I really enjoy driving, how can it be a good thing if that is taken way? Millions of professional drivers could be put out of work. I have a wife with mobility issues who would not be able to commute currently but self-driving vehicles could support her. It could make it economically viable to benefit people with mobility issues."	6	"Thinking about it in a more structured/scientific manner I have got concerns over it. There's lots of government support but not clear what the end game is: what do they want to see as the final outcome? I'm concerned that in London there will be widespread use of the buses which could ruin the current integrated transport system and mean a loss of jobs for professional drivers."	8
"More and more automation nowadays - we have to think less and less and we're doing away with jobs."	5	"I'm very sceptical about the safety of the technology. But if they are safer than what we have currently then maybe there is a future. I don't see how they could fit in to the current system though."	6
"I think it's a big issue there are many social, political factors. But I'm open to be convinced."	7	"it's about trust that things will work out well – I think it could free us all driving long journeys. but we're asking for too many things, that we trust that they can nudge through traffic, not hit people, we're being asked to trust that operators are acting in our best interests and I'm not sure."	7
Mean = 6.71 ; Range = 5		**Mean = 7.29; Range = 4**	

COMMENTARY FROM HUMANISING AUTONOMY

Overall, the project findings confirm what many experts within the industry state as well: the public is reluctant, but ready for SDVs. This reluctance is understandable and comes from a variety of sources – many of which are explored in this paper already. As a technical member of the wider SDV supply chain, it's important to highlight that through a collaborative and targeted approach, it will be possible to address many of the public's concerns in the short term, paving the way for future market acceptance and improved safety and overall mobility experiences for all.

How can we convince a sceptical public that they are trustworthy? What will that look like from a technology, data and/or regulatory perspective?

First off, the key here is that as a behavioural AI company, it is ours and the industry's responsibility to ensure this technology is responsibly used and promotes the safety and well-being of society. We understand and firmly believe that explainability, trust, accountability, and responsibility are guiding principles for any SDV, or generally advanced technology. While ethical AI is difficult to achieve; this study has once again proven it is essential that people's privacy, safety and responsible use of their data is considered in developing AI products.

Why is ethical AI so difficult to achieve?

Teaching morality to machines has never been an easy feat, and complex multi-system robotics make it even more difficult. Whilst current AI products often have simple goals, the future looks to widen the scope, creating products that use multiple AI features simultaneously. These products will combine data from a much wider set of sources, for several different aims.

For example, self-driving vehicles and smart cities will need to grapple with questions of new, abstract data, the goals of these systems, and how to safely deploy them. When considering ethics, safety and privacy, this data-fusion complexity brings new questions and challenges. AI systems need to be able to co-exist not only with other systems, but also with humans. Therefore, the communication of these systems is as important as the systems themselves.

Building Ethical AI Models

With these complications in mind, prioritising privacy, understanding the risks as a safety-critical function for autonomous systems, and the responsible usage of data can and need to be prioritised. For example, it's possible to extract key aspects of human actions from video footage in real-time, and these features are then used as a basis for high-level behaviour models. This provides privacy to our subjects, as the system does not divulge any personal information to the higher-level systems and focuses on communicating only the necessary attributes. Trustworthy AI design can mean that the AI can do all that it promises, without tracking your identity or recording your exact actions – it can be designed to respond to concrete and observable behaviours instead.

By embedding this modularity within systems, outputs can be broken down to their components, essentially building in aspects of explainability. This is part of a 'white box' approach (rather than 'black box') – to provide the ability of individuals to investigate the performance and inference of the system beyond inputs and outputs. A white box approach is also preferred when explaining the system to end-users and interfacing systems.

We, as a society, still have a long way to go until we see robots acting with true moral understanding, but the work in shaping safety and ethics standards starts now. With the introduction of 5G, new edge cases, and the rise of simulation-based validation and verification, there are opportunities to emphasize the need for not only safe, but also ethical products.

The challenge will be not to over-regulate the domain, but to align efforts and assure that legal requirements remain technically achievable and proportionate to assure that AI can be beneficial for society in the long run. This will require harmonised, data-driven regulations and it's going to require companies to be open with how they use data. Since AI can be so complex, especially when combined with machine learning and built from deep neural networks hidden within black boxes, it is especially difficult to understand where to begin – and to extract who should have accountability and ownership of the consequences when things inevitably go wrong.

As the world continues to become more automated, using ethical AI will reinforce human equality in our cities and urban mobility systems, the industry can move society towards a more sustainable, ethically conscientious, and inclusive world. People will trust AI more as they begin to understand how it can enable a better quality of life for them – and how the risks are minimised.

COMMENTARY FROM HUMANISING AUTONOMY

How can SDVs be programmed to evoke trust?

It's important to note that there will most likely be a natural, gradual increase in trust of SDVs as they become more mainstream, accessible, and are exposed to varying design domains and road traffic scenarios which will help improve performance. However, to speed up this process and to start with a higher trust bench line, it's imperative that manufacturers and regulators work together to mandate that supply chain components are built on ethical pillars to begin with. If the requirements for SDV components are increased, it would naturally mean that the end SDV system is a more trustworthy system.

Many of the workshop participants in this study most likely only thought about the SDV as a complete system, but it will be of vital importance to breakdown the many components and to understand how each individual part meets trust, ethics, and privacy requirements. An example of how a perception component can meet all these aspects to contribute to the overall SDV trustworthiness is described below.

Comparing AI / ML approaches to help define more ethical and trustworthy outputs

For an experienced driver, decision making at the wheel is intuitive. Drivers instinctively make thousands of micro-choices at the controls: when to adjust the steering wheel, choosing an efficient route and checking your mirrors all while following the rules of the road. Acceleration, deceleration, braking and steering all happen in the blink of an eye; avoiding collisions is second nature as the driver seamlessly predicts whether or not a person, animal or other object will collide with the car.

For human drivers, this becomes more difficult in more complex environments with larger numbers of actors to detect and track. While diverse pedestrian behaviour means human drivers have difficulty decoding what their counterparts on the road will do; SDVs will struggle even more – so what's the best approach to proceed with?

End-to-end deep learning is powerful, but has drawbacks

There's been some recent hype around using end-to-end deep learning for pedestrian crossing prediction, but it is not without its drawbacks. This approach uses large amounts of annotated video data that show diverse pedestrian behaviours. Now, we can predict crossing for pedestrians that have not been seen before by just looking at past and current behaviour represented in the pixels of the video.

This method is very powerful. However, end-to-end deep learning implies very few constraints on the structure of the model. There may be billions of parameters in these models. This complex structure means it is near impossible to understand how decisions are made; nor can we obtain reliable and valid estimates of prediction uncertainty. Deep learning models are known for their overconfident predictions (see Adversarial Examples). This black box approach makes end-to-end deep learning difficult to justify in safety critical applications like autonomous driving.

COMMENTARY FROM HUMANISING AUTONOMY

Physics models have interpretability, but lack complexity

Many state-of-the-art SDV and ADAS systems use deep learning to detect and localise pedestrians but need to rely on transparent models to predict pedestrian crossing. Physics models is a term used by Humanising Autonomy to describe the combination of noisy sensor data with short-term predictions that are formed by our knowledge about how objects propagate in the physical world. The models can be used to predict object locations for a set period of time.

By building models for the pedestrian and the vehicle separately and combining their output with map information, it is possible to compute the probability of a pedestrian crossing in front of the vehicle. We call this the physics model approach because it uses past location, their derivatives and infrastructure information to predict pedestrian crossing. In contrast to end-to-end deep learning, the developer imposes rigorous structure on the underlying model and relevant input.

This limits the number of parameters, distributions, and interactions between variables to consider. These restraints result in reduced model complexity, which makes a physics model approach more manageable than end-to-end deep learning. Although the physics model approach is widely accepted as the secure standard for SDVs and ADAS, our research shows that it is not clear if physics model prediction capabilities go far enough to enable a smooth, safe driving experience in pedestrian dense environments.

Moving beyond physics to tackle the limitations of camera perception

At Humanising Autonomy, we believe the missing link in pedestrian behaviour prediction is a true understanding of the pedestrians' cognitive processes. Pedestrians know when crossing is safe or not; drivers can intuit when he or she is unsure. This theory of mind helps to identify when further communication is required in road scenarios to establish a smooth interaction. Moreover, this perspective is necessary to bridge the critical safety gap in the industry's current approach to prediction models.

The Behaviour AI approach is not only more trustworthy and privacy preserving but is also proven to be safer. A quantitative analysis revealed that behavioural models can reduce the error of physics model crossing predictions by more than fifty per cent. In addition, it can predict crossing with an accuracy of 90% up to 4 seconds in advance; gradually increasing for shorter predictions (99% up to 1 second). In addition, these results surpass most deep learning approaches, but the behavioural model is far more transparent.

Without capturing the underlying psychological processes of pedestrian behaviour, physics models fail to accurately predict pedestrian crossing. Predictions can be wrong or delayed, which makes driving through downtown more difficult and potentially dangerous. By incorporating psychology into probabilistic machine learning models, it is possible to mitigate the limitations of physics-based models while keeping their positive attributes of a white box approach: interpretability, transparency, small model size and a trustworthy estimate of its prediction uncertainty.

CONCLUSION AND DISCUSSION

The aim of this project was to examine how to engage with communities to try to extract the values they would prioritise when it comes to the deployment and behaviour of **self-driving vehicles** (SDVs). This prioritisation of values could form the basis of an **ethical goal function** (EGF), which it is suggested may be an essential element for developing trustworthy, predictable and safe SDV behaviour. Furthermore, the development of an EGF for SDVs would help **developers** and **regulators** ensure the technology meets the needs and expectations of the communities into which they are deployed.

Since the potential applications for SDVs are so broad, we focused our study on a use case that has been frequently suggested as one of the candidates for early implementation – that of a small **self-driving bus** operating as a transport service in a city environment (such vehicles are being developed by e.g. ZF, EasyMile etc.).

The results of the survey and workshops revealed significant **variation** of views around SDV behaviour, highlighting the challenge in trying to extract values that reflect the overall expectations and values of a community. The emerging priority for the public was for behaviours and operation of SDVs that build trust in their use.

This included a preference for vehicles to **comply** strictly with existing regulations and proof that, as a minimum, the operation of SDVs **did not increase risk of harm** for occupants or other road users. There was an understanding that collisions involving SDVs may still occur but there was common agreement that developers should not cut costs at the expense of safety.

There was also a strong preference that SDV operators should be **transparent** with data, **sharing** relevant information with government investigators to enable lessons to be learned around future safety. Sharing of operational data between SDV companies in the interests of safety was also raised as being potentially useful.

Results indicated that SDVs should seek to preserve the safety of vehicle occupants and other road users **equally** and that the **interior design** of the vehicles should account for the possibility that the vehicle may need to swerve or brake sharply to prevent an incident with another road user (e.g. with grab handles, soft surfaces etc.). Compared to other values, participants were seemingly quite relaxed about a possible need to change the design of urban infrastructure to accommodate SDVs – although participants may have been less indifferent if they had more information about what that might entail (for example, pedestrian barriers, dedicated SDV lanes etc.).

For many topics there was significant **diversity** of opinion, highlighting that it would be difficult to define EGFs that cause vehicles to behave in line with the expectations of **all** citizens - and that the behaviours SDVs adopt in their early deployments are likely to encounter resistance from some members of the public.

We identified several '**ethical red lines**' for which there seemed to be broad support – and which could therefore be considered as a starting point for the EGF:

- Safety should be improved.
- Operation should be within a clear legal framework.
- Taking risks to save time or cost is not acceptable.
- All road users should be protected equally.
- Data sharing with reasonable stakeholders (e.g. insurers, police etc.) is acceptable, provided this is done transparently and without conflicting with data privacy and data protection regulations.

CONCLUSION AND DISCUSSION

Further work is required to refine these. For example, it is not clear what level of safety improvement would be classed as acceptable or how you would establish a suitable pre-deployment baseline for comparison with safety performance post deployment. Similarly, risk could be reduced by driving very slowly but this is not practical for an efficient transport service so to achieve safety, an SDV must be able to navigate its environment successfully, making accurate predictions about the future behaviour of other road users when necessary and operate within infrastructure that promotes safe and appropriate interactions. Research is therefore needed to explore how factors such as SDV type, use case, operational design domain, location, community affected and so on influence the definition of the red lines.

Responses appeared broadly similar across the age range for our survey respondents. However, positivity towards SDVs declined with age until we reached the two groups over the age of 65 where positivity seems to increase slightly. It is suggested that this reflects a recognition from participants that SDVs could support independent mobility later in life and so they are more upbeat about their use – but this needs further research.

An intriguing inconsistency emerged between the survey and workshop results. Survey respondents were very positive towards the statement suggesting that an SDV **should** break the rules to avoid a collision (using the example of crossing a double white line to avoid a pedestrian). This resulted in category of 'legality' being rated as one of the least important. However, workshop participants felt compliance with traffic rules was **essential** and it was rated as one of the most important of the eight ethical dimensions they were required to rank.

This highlights the need to explore further the extent to which participants **expect** strict compliance with road rules, identifying if there are situations where non-compliance may be preferable and determining whether SDVs could reliably determine when and how non-compliance would be optimal.

Engaging with communities to develop of EGFs in such a way that is useful for SDV developers is challenging. An alternative solution is to leave the responsibility for ethical decision-making with SDV users. Contissa, Lagioia and Sartor (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 suggested 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. As Reed et al. (2021) surmised, this approach places considerable responsibility on the user with the 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 SDV in the three modes on the technology developers, which again may not produce outcomes seen as socially or ethically acceptable. With participants in this study clear that trust was an essential component of their acceptance of SDVs, it seems that engaging with communities to ensure that the technology aligns with their preferences is a more productive approach.

The results highlight that a **matrix** of elements must coalesce to achieve successful SDV deployment. SDVs must operate in a **trustworthy** manner. This means that developers must understand what it means for road users and the wider public for SDVs to be 'trusted' – with **EGFs** a route to achieving that. It also means that **regulations** should be in place to ensure that SDVs operate **legally** and so that, when collisions occur, the expectation that we **learn** from incidents is met.

This learning might be in the form of updates to regulations, it could be a requirement for an SDV developer / operator to withdraw and update SDV hardware or software, it could be for a road infrastructure authority to change the way in which its roads are designed, maintained or operated. As responsibility for an incident is understood, a suitable **insurance** process should make sure that those affected are adequately compensated for any loss.

The breadth of these elements emphasises the need for an **ecosystem approach** to transport innovation with innovators, regulators, political leaders, infrastructure providers, insurers and the public all playing essential roles.

CONCLUSION AND DISCUSSION

To improve the sensitivity of the development of the EGF, it is suggested that public surveys and workshops are developed using **simulated scenarios** (rather than verbally described or text-based) that enable participants to gain greater context on a variety of SDV behaviours. Whilst it is difficult to determine the extent to which survey participants understood the scenarios described in each question, it was clear that workshop participants often had different interpretations and understanding of the SDV scenarios being described to them.

Consequently, differences in opinion may reflect this variation in understanding rather than genuine discrepancies in values or preferences. Computer generated simulations of scenarios could provide a much clearer representation of ethically ambiguous scenarios, allowing participants to give better informed responses to those scenarios. Furthermore, the simulations could be **tuned** such that the parameters of the scenario are dynamically adjusted according to the responses of participants – for example, if one respondent thinks that it was safe for the SDV to pull out into a gap in traffic, the next respondent might see a modified scenario where the gap is smaller.

Scenarios could be continuously refined until an **optimal** answer is achieved for the community under test. This would enable ethical factors to be tested and compared to tune the parameters that underlie EGFs. Ultimately, this then determines the behaviour of SDVs in accordance with the requirements of the community or user group affected by the deployment of the SDVs. The results of such a study would support:

- SDV **developers** – in providing objective evidence against which they could tune the performance of their systems such that they align with the expectations of the communities into which they are to be deployed.
- SDV **regulators** – in setting the parameters (i.e. the EGF) for safe operation of SDVs.
- The **public** – giving confidence that SDVs will behave in a safe and predictable manner.

The project has highlighted that it is possible to explore societal acceptance of SDV behaviours in ethically ambiguous scenarios and to seek consensus on how SDVs should behave. For many topics, the diversity of views suggests that agreement over a mathematical description of the how SDVs should behave – as required for the creation of an EGF – is not possible to achieve. However, there is cause for optimism. Firstly, the value participants rated as most important was that of ‘trust’. In relation to SDV operation, this may come in many forms; for example, trusting that...

- an SDV will behave as expected in a **dilemma** situation;
- an SDV will comply with existing **traffic rules**;
- an SDV will **protect** other road users as well as its occupants;
- SDV manufacturers will design and manufacture **safe vehicles**;
- regulators will be able to **access data** about SDV safety performance to ensure that the industry **learns** from any incidents that occur.

In terms of developing SDVs that will be accepted by society, the industry therefore needs to:

- Gain greater depth on how to build trust through **research** and **engagement**.
- Focus early SDV trials and deployments on use cases that maximise opportunities to **build trust** (and avoid those where there could be ambiguities about how the SDV should behave).
- Engage with communities that could be affected by SDV deployment and determine their desires, concerns or ‘**red lines**’ about SDV operation.
- **Communicate** how SDVs have been designed, tested and deployed in line with these requirements.

Technology regulators can support this process by providing guidance on what those seeking to trial or deploy SDVs should (and should not) do to gain the **confidence** of the communities where they operate. The results of the study also highlight that the public expects ongoing **dialogue** between developers, regulators and society to help evolve SDV behaviours towards those that are most acceptable.

The aim of the project was to use the EGF approach as a means by which that dialogue could progress and, whilst it has been difficult to extract a mathematical description of our participants’ values, it has highlighted how societal input can help us to understand how SDVs can and should behave.

CONCLUSION AND DISCUSSION

The response from **Humanising Autonomy** on the results of the study highlight how their approach to system development already recognises the importance of some of the ethical red lines and the need for users to build trust in the operation of a complex system. They recognise the importance of **transparency** of operation to support investigation of system behaviour and explainability.

They also note the potential risk of **over-regulation** – constraining innovation by defining system performance or architectures too soon – but suggest there needs to be **alignment** and **collaboration** between regulators and developers in making sure legal frameworks are **proportionate** while delivering public safety and societal benefit. A very interesting comment from Humanising Autonomy that did not emerge within the survey responses or workshops was their assertion that ethical requirements should cascade down into the **supply chain** and not just apply to the complete system.

With SDV manufacturers bringing together components from a variety of providers (as is typical in the automotive context), this creates an opportunity for these tier 1 and tier 2 suppliers to **differentiate** their products through factors such as trust, ethics and privacy in their design. Finally, Humanising Autonomy recognised the limitation of **black box** approaches in safety critical applications, echoing the assertion that such systems cannot derive ethical values – and that these must therefore be developed explicitly, which is what we have attempted to do for SDVs in this project.

One assumption inherent to our approach is that people genuinely want SDVs. We are aware that this might not be the case and certainly coming into the workshops some participants were highly sceptical of the real value of SDV technology. However, whilst the media attention and investment associated with a new technology is not always an indication of society's desire to adopt it (e.g. Segway, Google Glass etc.), it is probably fair to assume that SDVs in some form are likely to be adopted.

With that in mind, it is important that **societal interests** are considered in the way SDVs are designed, developed and deployed to ensure they are welcomed as a useful, beneficial (or at least acceptable) technology and avoid some of the problems that we have seen emerge with car dependence. This can be supported by regulations that guide SDV developers towards approaches that account for such interests, recognising that these can and will evolve over time and that will be different according to the location and use case of SDV deployment.

The project has begun to elucidate a process and structure by which regulators can **engage** with the public in the acceptable and desirable features of SDVs. It is an approach that is not just applicable to SDV technology but also to other AI-based transport innovations that may emerge – and to help maximise the probability that they are welcomed and embraced by an informed majority based on **robust evidence** and not scuppered by a minority of vocal opponents citing anecdotes and opinions. In the coming fifty years, we believe that this will be vital if we are to fulfil the vision of **Rees Jeffreys** and ensure that our roads genuinely deliver **prosperity** and **enjoyment** for future generations.

REFERENCES

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). IEEE.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F. & Rahwan, I. (2018). The moral machine experiment. Nature, 563(7729), 59-64.

BSI (2022) CAV Vocabulary. BSI Flex 1890 v4.0. Accessed April 2023 click here.

Coalition For Future Mobility (2017). Highly automated technologies, often called self-driving cars, promise a range of potential benefits. Accessed April 2023 click here.

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.

Cruise (2022). Driving cities forward. Accessed April 2023 click here.

Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology). Accessed April 2023 click here.

Department for Transport (2022). Transport and environment statistics 2022. Accessed April 2023 click here.

Etienne, H. (2021). The dark side of the 'Moral Machine' and the fallacy of computational ethical decision-making for autonomous vehicles. Law, Innovation and Technology, 13(1), 85-107.

European Parliament (2020) 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. Accessed April 2023 click here.

Foot, P. (1967) "The problem of abortion and the doctrine of double effect." Oxford Review 5

Hickman, R. (2019). Urban growth, the private car and inactivity. Town & Country Planning, 88(1), 4-8.

HM Government (2022). Connected & Automated Mobility 2025: Realising the benefits of self-driving vehicles in the UK. Crown copyright 2022, OGL v3.0. Accessed April 2023 click here.

Jung, J. C., & Sharon, E. (2019). The Volkswagen emissions scandal and its aftermath. Global Business and Organizational Excellence, 38(4), 6-15.

Law Commissions (2022) Automated Vehicles: joint report. Law Commission of England and Wales, Law Commission No 404. Scottish Law Commission, Scottish Law Commission No 258. Accessed April 2023 click here.

Mattioli, G. (2021). Mattioli, G. (2021). Transport poverty and car dependence: A European perspective. In Advances in Transport Policy and Planning (Vol. 8, pp. 101-133). Academic Press.

Reed, N., Leiman, T., Palade, P., Martens, M. H., & Kester, L. (2021). Ethics of automated vehicles: breaking traffic rules for road safety. Ethics and Information Technology. Accessed April 2023 click here.

Stilgoe, J., & Cohen, T. (2021). Rejecting acceptance: learning from public dialogue on self-driving vehicles. Science and Public Policy, 48(6), 849-859.

Waymo (2018). Safety. Accessed April 2023 click here.

Zimmer, J. (2016). The Third Transportation Revolution – Lyft's Vision for the Next Ten Years and Beyond. Accessed April 2023 click here.

APPENDIX A. USE CASE INFORMATION

The following information was provided to survey and workshop participants to explain the SDV use case that they should think about when responding to questions.

Imagine that last year, your local council introduced a public transport service from an operator called AutoCityBus that uses self-driving buses.

These buses can take up to 12 passengers at a time (see below).

[Example image of Gacha SDV shuttle bus provided]

The self-driving buses:

- drive up to a maximum speed of 25mph;
- only operate on roads with a speed limit of 30mph or less;
- operate without a human driver in the vehicle;
- have cameras to monitor passenger welfare;
- are connected to a control centre that can help if there are any problems (e.g. vehicle breakdown);
- operate in an area within 5 miles of the main train station.
- have passed a government certification scheme that indicates they are adequately safe to use as part of the AutoCityBus service.

The service is designed to operate in an 'on demand' manner.

For example, local resident, Sandeep, uses the bus to commute to his nearest train station each morning and home again each evening. He uses a smartphone app to call for a bus when he wants to leave and to say where he wants to go. The app then tells Sandeep to go to a collection point on the main road a short walk from his home address when the bus will arrive to collect him.

The full commute usually takes about 15 minutes but the route and the journey time may vary from day to day depending on how many other passengers call for the bus and where they wish to go. However, the service offers passengers a guarantee that their journey will never be delayed by more than 5 minutes, with refunds available if this is exceeded.

APPENDIX B. SURVEY STATEMENTS

These are the 30 statements against which participants were asked to provide their level of agreement:

1) I would be happy for the self-driving bus to take more risks (e.g. pulling out into a smaller gap between traffic at a junction) to catch up time if it had been delayed in a traffic jam.
2) I would want the self-driving bus to take emergency action to avoid a pedestrian who unexpectedly stepped in the road, even if it meant risking injury to passengers on board the vehicle.
3) The self-driving bus should behave to suit the roads where it is operating (e.g. pulling out more assertively at busy town junctions compared to quieter country roads).
4) If the self-driving bus has had to wait a long time to pull out at a junction and a queue of traffic has formed behind it, it should carefully edge out into traffic to encourage other road users to give way.
5) The self-driving bus should leave a wider gap to pedestrians and cyclists than when passing stationary vehicles, even if that means putting its passengers at a slightly greater risk of collision with oncoming traffic.
6) If the self-driving bus faces an unavoidable collision, the life of a family pet is just as important as that of a human family member.
7) It is acceptable for the self-driving buses to crash sometimes, as long as they are safer overall than human controlled vehicles.
8) If a collision with another vehicle is unavoidable, the self-driving bus should try to protect its passengers as its top priority, even at the expense of other road users.
9) The self-driving bus should not break the rules of the road to avoid holding up traffic (e.g. crossing double white lines to overtake a cyclist who is travelling at 14mph).
10) The self-driving bus should break the rules of the road if it means avoiding a collision (e.g. swerving across double white lines to avoid a pedestrian who has stepped into the road).
11) The self-driving bus should move out of the way of an ambulance attending an emergency, even if this means the self-driving bus has to break the rules of the road (e.g. driving through a red traffic light to give space for the ambulance to pass).
12) The self-driving buses should drive at speeds that keep up with the traffic flow (within the speed limit), even if this increases risk to pedestrians.
13) The self-driving bus should have lights or markings to show that there is no human controlling the vehicle.
14) Schools should encourage parents to send their children to school in a self-driving school bus to reduce congestion caused by ‘school run’ traffic.
15) It would be worth taking road space from other traffic in large towns and cities to give self-driving buses their own lane so it is easier for them to get around.

APPENDIX B. SURVEY STATEMENTS

16) Cities should put up barriers on pavements to make it easier for self-driving buses to drive through busy areas.

17) Self-driving buses would improve city centres because people would switch to using the buses rather than driving into the city.

18) I trust that the UK government will require self-driving buses to be suitably tested before they are allowed on public roads to ensure that they operate safely.

19) I would trust technology companies (e.g. Google, Apple) to produce self-driving buses that operate safely on UK roads.

20) I would trust vehicle manufacturers (e.g. Mercedes-Benz, Volvo) to produce self-driving buses that operate safely on UK roads.

21) The UK government should not be heavy-handed over safety regulations that might delay the growth of the industry.

22) The data collected by a self-driving bus should remain the private intellectual property (trade secrets) of the manufacturer so long as crash risk is reduced compared to human driven vehicles.

23) The data collected by a self-driving bus must be shared with government investigators in the event of a crash to help understand why it happened and who was to blame.

24) Self-driving bus operators should be required by law to keep a record of their vehicles' actions in the interests of safety, even if it is expensive to store all the data.

25) I would accept the private companies operating self-driving bus services using my personal data on the journeys I have taken if it means safer roads.

26) There will need to be continuous surveillance of passengers on a self-driving bus in the interests of safety.

27) A passenger on a self-driving bus must always be able to stop the vehicle using an emergency button in the passenger compartment.

28) Road deaths and serious injuries are an inevitable price to pay for the convenience and benefits of motor vehicles.

29) On the whole self-driving technologies (e.g. cars, buses) are a good thing.

30) I would use a self-driving bus service tomorrow if it were available to me.

APPENDIX C.
SURVEY STATEMENT CATEGORY WEIGHTINGS

Statement	Mobility	Legality	Trust	Safety of passengers	Safety of others	Cost	Fairness	Urban design
I would be happy for the self-driving bus to take more risks (e.g. pulling out into a smaller gap between traffic at a junction) to catch up time if it had been delayed in a traffic jam.	1		1	-1	-1	1		
I would want the self-driving bus to take emergency action to avoid a pedestrian who unexpectedly stepped in the road, even if it meant risking injury to passengers on board the vehicle.				-1	1			
The self-driving bus should behave to suit the roads where it is operating (e.g. pulling out more assertively at busy town junctions compared to quieter country roads).	1			-1	-1			
If the self-driving bus has had to wait a long time to pull out at a junction and a queue of traffic has formed behind it, it should carefully edge out into traffic to encourage other road users to give way.			1	-1	1		1	
The self-driving bus should leave a wider gap to pedestrians and cyclists than when passing stationary vehicles, even if that means putting its passengers at a slightly greater risk of collision with oncoming traffic.					-1		1	
If the self-driving bus faces an unavoidable collision, the life of a family pet is just as important as that of a human family member.	-1	1						
It is acceptable for the self-driving buses to crash sometimes, as long as they are safer overall than human controlled vehicles.			-1	1				
If a collision with another vehicle is unavoidable, the self-driving bus should try to protect its passengers as its top priority, even at the expense of other road users.				1	-1			
The self-driving bus should not break the rules of the road to avoid holding up traffic (e.g. crossing double white lines to overtake a cyclist who is travelling at 14mph).		1		-1				
The self-driving bus should break the rules of the road if it means avoiding a collision (e.g. swerving across double white lines to avoid a pedestrian who has stepped into the road).		-1		1	1			
The self-driving bus should move out of the way of an ambulance attending an emergency, even if this means the self-driving bus has to break the rules of the road (e.g. driving through a red traffic light to give space for the ambulance to pass).	1	-1			-1			
The self-driving buses should drive at speeds that keep up with the traffic flow (within the speed limit), even if this increases risk to pedestrians.				1			-1	
The self-driving bus should have lights or markings to show that there is no human controlling the vehicle.			1					-1
Schools should encourage parents to send their children to school in a self-driving school bus to reduce congestion caused by 'school run' traffic.			1	1				
It would be worth taking road space from other traffic in large towns and cities to give self-driving buses their own lane so it is easier for them to get around.				1	1			-1
Cities should put up barriers on pavements to make it easier for self-driving buses to drive through busy areas.				1	1	-1		-1
Self-driving buses would improve city centres because people would switch to using the buses rather than driving into the city.								1
I trust that the UK government will require self-driving buses to be suitably tested before they are allowed on public roads to ensure that they operate safely.			1					
I would trust technology companies (e.g. Google, Apple) to produce self-driving buses that operate safely on UK roads.			1					
I would trust vehicle manufacturers (e.g. Mercedes-Benz, Volvo) to produce self-driving buses that operate safely on UK roads.			1					
The UK government should not be heavy-handed over safety regulations that might delay the growth of the industry.	1		-1					
The data collected by a self-driving bus should remain the private intellectual property (trade secrets) of the manufacturer so long as crash risk is reduced compared to human driven vehicles.			-1			1		
The data collected by a self-driving bus must be shared with government investigators in the event of a crash to help understand why it happened and who was to blame.			1			-1		
Self-driving bus operators should be required by law to keep a record of their vehicles' actions in the interests of safety, even if it is expensive to store all the data.			1			-1		
I would accept the private companies operating self-driving bus services using my personal data on the journeys I have taken if it means safer roads.			-1	1	1			
There will need to be continuous surveillance of passengers on a self-driving bus in the interests of safety.			1	1		-1		
A passenger on a self-driving bus must always be able to stop the vehicle using an emergency button in the passenger compartment.			1					
Road deaths and serious injuries are an inevitable price to pay for the convenience and benefits of motor vehicles.	1			-1	-1		1	