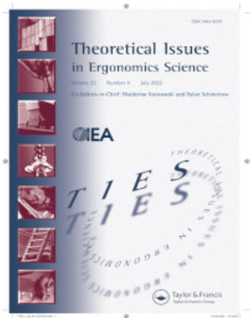

Theoretical Issues in Ergonomics Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttie20

Ethical considerations and moral implications of autonomous vehicles and unavoidable collisions

Jonathan Robinson, Joseph Smyth, Roger Woodman & Valentina Donzella

To cite this article: Jonathan Robinson, Joseph Smyth, Roger Woodman & Valentina Donzella (2022) Ethical considerations and moral implications of autonomous vehicles and unavoidable collisions, Theoretical Issues in Ergonomics Science, 23:4, 435-452, DOI: 10.1080/1463922X.2021.1978013

To link to this article: https://doi.org/10.1080/1463922X.2021.1978013

Published online: 15 Oct 2021.

THEORETICAL ISSUES IN ERGONOMICS SCIENCE
2022, VOL. 23, NO. 4, 435–452
https://doi.org/10.1080/1463922X.2021.1978013

Ethical considerations and moral implications of autonomous vehicles and unavoidable collisions

Jonathan Robinson, Joseph Smyth, Roger Woodman and Valentina Donzella

WMG, University of Warwick, Coventry, UK

ABSTRACT
While it is widely agreed that automated and autonomous vehicles may provide safety benefits over vehicles with lower level or no automation, due to other road users there will still likely be situations where a collision is unavoidable. What should a vehicle that is operating autonomously do when it has no choice but to have a collision? And who should decide which vehicle manoeuvre is the most acceptable? These situations create moral dilemmas requiring consideration of the most acceptable and moral action of the vehicle. In this paper we explore current research in this domain and work towards enabling ethical solutions. We identify current experimental work (practical studies rather than theoretical studies) on this problem often contains fundamental flaws due to the lack of real-world validity within the studied scenarios. We argue that morality is highly context dependent and that participants need to be more engaged in the choices they are claiming to make. Suggestions for future work include virtual reality or simulation methodologies which promote immersivity to ensure procedural validity whilst retaining safety. We also identify current guidelines contradict public viewpoint and argue public attitude needs to be better understood to give autonomous and automated vehicle manufacturers confidence in their design.

ARTICLE HISTORY
Received 28 January 2021
Accepted 3 September 2021

KEYWORDS
Autonomous vehicles; trust and acceptance; moral machine; ethics; unavoidable collisions; safety

Relevance to human factors

If AVs are going to make their way onto our public roads, they must satisfy the demands, expectations and desires of those interacting with them, using them or sharing space with them (Awad et al. 2018; Lin 2016). Hence, future AVs must be accepted by the public to be successfully integrated into our society. How they behave when they have no choice but to have a collision needs to stand up to any scrutiny posed by members of society. Therefore, this problem is highly human related. A decision that would usually be made by a human in a short span of time is replaced by the AV's decision-making process, by a directly or indirectly pre-designed decision made over a longer period of time. Thus, public acceptance of this decision is of foremost importance.

CONTACT Joseph Smyth j.smyth.4@warwick.ac.uk WMG, University of Warwick, Coventry, UK.

Introduction

With the introduction of vehicles with greater levels of automation we could see vast improvements to parts of everyday life. Not having to drive a vehicle could free up vast amounts of time, allowing more productivity (and potentially shorter office hours) (Smyth, Jennings, and Birrell 2020), increased autonomy for those unable to drive through disability or otherwise, and increased safety for road users (Fagnant and Kockelman 2015; Anderson et al. 2014). However to fully realise these benefits, as with any new technology, certain barriers must be overcome. The technology of automated vehicles (AVs) must be accepted by the public if they are to be the consumers of the technology (Adnan et al. 2018; Shariff, Bonnefon, and Rahwan 2017). Improving public perception of AVs can increase the likelihood of acceptance (Haboucha, Ishaq, and Shiftan 2017). However, this concept of acceptance is only really explored in this pre-AV world as until March 2021, when Honda released the Honda Legend as the first Conditionally Automated Vehicle (Honda European Media Newsroom 2021), there were no vehicles with this technology on the roads. The Society of Automotive Engineers (SAE) have provided a very useful classification of the levels of automation within vehicles, with Level 0 denoting 'no automation' and Level 5 being 'full automation' (SAE 2021). It is Level 4 and 5 that have the potential to bring maximum societal benefits – but these levels are also the most complex considering technological implementation but also public acceptance. At Level 5 the vehicle is in full control of the complete dynamic driving task, including its longitudinal and lateral control and all sensing and perception of the environment. The driver should never be expected to intervene or take control of the vehicle (other than perhaps entering a destination). A Level 4 vehicle boasts similar technological abilities within specified occupational design domains, although the driver can choose to take over control at any point (within as yet undefined transition periods) if they wish. Level 3 vehicles will also be considered as at times the vehicle is entirely functioning without driver engagement under its conditionally automated system. Previous work by Flemisch et al. has highlighted that in vehicles with automation, authority and responsibility with respect to the control of the vehicle is shared and the proportion of which it is shared is determined by the ability of the human driver and vehicle to perform an action in a given state (Flemisch et al. 2012). In this paper, we are considering the situation where a vehicle is unable to transfer control back to the human driver and hence has authority and control over the actions. A vehicle under these circumstances may not have the ability to avoid a collision but does have the ability to choose between actions whilst not necessarily being able to determine the responsibility for consequences of the actions.

Recent research suggests that 'pre-acceptance' of such vehicles is already strong, with much of the public excited about the technology (Payre, Birrell, and Parkes 2021). However, we must also consider how to maintain this acceptance when these vehicles are on the road. A common question from the public related to the acceptance of future vehicles with higher levels of automation is often discussed around the vehicles' actions in case of a collision (Awad et al. 2018; Bonnefon, Shariff, and Rahwan 2016). Unfortunately, AV collisions will happen like the Uber collision in Arizona and these can attract large amounts of media attention, so appropriate actions in these events must be investigated (Levin and Wong 2018; Lee 2019; BBC 2020; Conger 2020; McCausland 2019). Let's consider a hypothetical (but possible) scenario whereby an AV operating under its automated driving system is

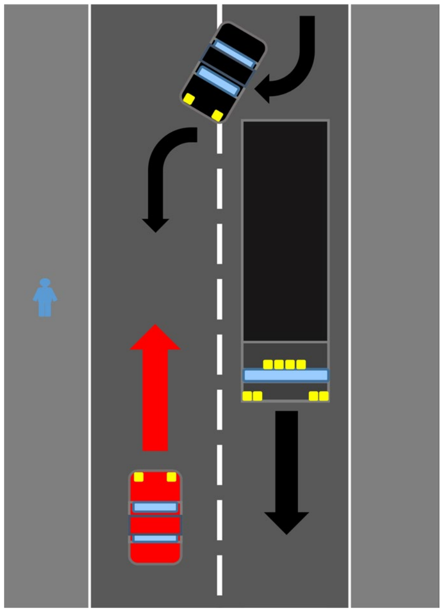

Figure 1. Visual representation of an unavoidable collision.

driving along a single carriageway road in a pedestrianised area when suddenly a human driven vehicle travelling in the opposite direction swerves into the AV's lane. The AV identifies the upcoming hazard and applies the brakes, but due to the circumstances outside of the vehicles control, a collision is inevitable. What should the vehicle do? It could stay on its current trajectory and collide with the oncoming vehicle. It could swerve into the opposite lane and risk colliding with the adjacent oncoming vehicle. Or it could swerve towards the pavement and risk colliding with a pedestrian (Figure 1).

This example forms a moral dilemma whereby the AV's automated system is forced to make a decision that will have significant implications no matter the outcome. If the vehicle stays on its current path, the oncoming vehicle that caused the accident bares all responsibility for the collision but the head-on collision could be more severe for the occupant of the AV. Since those buying an AV could well be doing so for safety reasons, this could create a conflict of interest between the consumer desire for safety and the manufacturer's liability concerns. If the vehicle collides with the pedestrian, the occupant of the AV will be safer, but the pedestrian will be in great danger. Colliding with the other vehicle would perhaps be less dangerous than a head-on collision but means involving a vehicle (and its occupants) in a situation that they had no part in causing. This is just a hypothetical example of a potential unavoidable collision, but already contributing factors start to become apparent. How might this decision change if there were children involved, for example (as occupants of the vehicles or pedestrians)? There are endless hypothetical examples and contributing factors for such a scenario and it is beyond likely that future vehicles will inevitably encounter such issues and have to deal with the consequences.

In order to maintain public acceptance and trust in these future technologies we must consider now, before such an event occurs, what is the 'right' thing to do. Even beyond acceptance, technology should be designed to benefit society, and we must consider how this can be achieved. In this paper, we will discuss the ethics of what a moral action may look like in an unavoidable collision for a vehicle capable of autonomous driving. In the context above, there are multiple possible solutions, and the perceived correct action may differ between individuals. We explore the current literature and demonstrate the complexity of the factors within such a decision-making process – bringing to the forefront some of the most important factors that the public consider in determining the ethical solution. We show an appropriate solution will require the input of many different stakeholders including vehicle manufacturers who will want to understand how to design their vehicles and legislators who will want to ensure that the vehicles act appropriately to benefit society. We show how this solution is context dependent and how appropriate it is may vary with country or culture. However such decisions cannot be made by regulators or vehicle manufacturers alone. Members of the public will want to ensure the vehicles meet their expectations, both as a potential AV owner and as a member of society that shares space with AVs.

Trolley vs transplant problem

When discussing what an AV should do when it has no choice but to have a collision, people often consider the famous thought experiment referred to as the 'Trolley Problem', devised by Philippa Foot in 1967 (Gorr 1990). This is a hypothetical thought experiment whereby a runaway mining trolley is heading down a track towards five workers who will be killed on impact. However, there is a bystander who can switch the tracks such that the trolley will bypass the five and instead now collide with and kill one worker, saving the original five. According to Gorr (1990), most people in this situation choose to switch the tracks, saving the five at the expense of the one. An alternative 'problem' with similar results is that of the hypothetical 'Transplant Problem', where a surgeon can save the lives of five patients by performing organ transplants using organs from one healthy living person who would have otherwise lived (Gorr 1990). In this situation, most people say that the surgeon would be immoral to perform the transplants (Gorr 1990) despite the same 'one life to save five' theme. If it were to be considered immoral to save five lives in the Transplant Problem but moral in the Trolley Problem, it becomes clear how context is extremely important when considering morality. These problems at their core are equivalent: save five lives at the expense of one and the disparity in solutions is directly transferable to AV ethics. Whilst it is acknowledged that thought experiments such as the Trolley problem and Transplant problem are hypothetical, they are useful for understanding how morality is dependent on context. Here we have two problems which are objectively equivalent but have apparently different most moral actions.

AVs might be faced by a moral decision in rare but not impossible situations where they cannot avoid a collision (Martínez-Díaz and Soriguera 2018; European Commission 2020). Whether the AV should choose who lives in such a situation or not is up for debate but what is clear is that a coherent, morally sound policy is needed for such a situation (Dimitrakopoulos, Uden, and Varlamis 2020). What the Trolley and Transplant problem show is that slight differences in context of the unavoidable collision can cause differences in the perceived most moral action. Hence, a decision may not be able to be based solely on facts such as 'number of lives saved'. As is often highlighted, the overall aim is to avoid these situations entirely, negating the need for this moral problem, perhaps just maintaining lane position and applying maximum deceleration when this is not possible (Burton et al. 2020; Luetge 2017; Lin 2016). However, choosing to just apply the brakes alone is in itself a policy, which therefore needs to stand up to scrutiny to prove its morality. If this policy meant colliding with a child in the middle of the road instead of swerving into wall, would this be considered morally sound? Questions such as these must be answered prior to deployment of vehicles that operate without the intervention of a human driver. If a solution is implemented that is not deemed socially acceptable post-deployment, trust in AVs could be significantly damaged (Dimitrakopoulos, Uden, and Varlamis 2020), not to mention the potential safety implications involved.

Current ethical guidelines

While laws and ethics are two different things (Robles Carrillo 2020), discussions around ethics such as this one can be used to inform the introduction of laws and policies (Schiff et al. 2020). Ethics become particularly important where there is little regulation or laws surrounding an area (Robles Carrillo 2020). Carrillo talks of a universal international organisation needed for the establishing of rights and commitments of the use of artificial intelligence (AI) in general (Robles Carrillo 2020) which stands as good evidence to the connected nature of these two areas. In this paper, we wish to demonstrate what is needed before a legal framework for AVs in unavoidable collisions can be developed. Leslie talks of artificial intelligence ethics being related to widely accepted standards of right and wrong (Leslie 2019). However, what is widely accepted as right and wrong is as yet unclear.

In response to this unavoidable collision problem, certain ethical guidelines have been devised to help build a framework for how an AV should react in an unavoidable collision scenario for example a report by the European Commission (European Commission 2020) and the German Ethics Code for Automated and Connected Driving (Luetge 2017). The vehicle must meet the state of the art for safety in collision avoidance and provide measures to prevent unsafe use by the users (European Commission 2020; Luetge 2017). This emphasises the need to reduce the probability of an unavoidable collision occurring, both by improving the ability of the AV to avoid collisions and ensuring that they are used appropriately by the users. When discussing moral dilemmas from unavoidable collisions it is important to remember that the more effective the collision avoidance is, the less important the moral dilemma becomes.

The German Ethics Code for Automated and Connected Driving (2017) recommends that no decision to deliberately choose one life over another should be taken by the programmer (Luetge 2017). Making it clear that no decision should be made to program a vehicle to preference one entity over another, such as choosing to save a child over an adult for example. The European Commission's Independent expert report recommends that AVs should redress existing inequalities in vulnerable road users to equalise the 'harm-relative-to-road-exposure' ratio for road users (European Commission 2020). It is important to ensure that this recommendation is only applied during normal operation and not during the policy in an unavoidable collision. If not, we could arrive at a situation whereby AVs are discriminating road users based on mode of transport. If this was the case, road users may be effectively 'punished' for using a safer mode of transport or where vehicles are designed/purchased that are less physically safe so they are effectively safer (Adnan et al.

2018; Mordue, Yeung, and Wu 2020; Lin 2016). This consideration is reiterated by a further ethical guideline from the German Ethics code recommending against any decisions being made based upon personal characteristics, emphasising the need to avoid discrimination of any form (Luetge 2017). Further to this, it is recommended that non-involved parties should not be sacrificed in a collision scenario (Luetge 2017). However there is significant ambiguity in the determination of who is 'involved'. Could a pedestrian on the pavement be involved? Or do they only become involved once the decision to swerve off the road is made? One analysis of the Trolley problem is that the bystander should not switch the tracks, as the one worker only becomes involved in the dilemma once the decision is made to switch (Gorr 1990). By similar logic the vehicle should not be programmed to make a decision in this scenario and should just collide with what is in front of it. However, would this be considered acceptable by the public and would AV customers still wish to buy these vehicles. If this guideline was to be followed, the involved parties would need to be clarified.

Humans are also recommended to be given priority over animals and property in the event of an unavoidable collision (Luetge 2017; European Commission 2020) – this is less ambiguous. However it is noted that caution is needed when considering cultural differences with regards to 'higher animals' (Luetge 2017), and further, collision with a large animal such as a moose will cause significantly different safety outcomes than, for example, a rabbit. The European commission report recommend that moral dilemmas should be managed 'by principles of risk distribution and shared ethical principles' (European Commission 2020). As the previous example of the trolley problem suggests though, do we have shared ethical principles? If so, what are they? This highlights a clear need to involve the public in the process of defining this to ensure that these shared ethical principles are in fact 'shared'. The European Commission report states that

> the behaviour of a CAV in a dilemma situation is by default acceptable if the CAV has, during the full sequence that led to the crash, complied with all major ethical and legal principles define in this report

(CAV – Connected and Automated Vehicle) (European Commission 2020). However, one of these ethical principles is 'inclusive deliberation' (European Commission 2020). This guideline can only be met if an accurate understanding of public viewpoint is obtained, so we must seek to explore this.

The unavoidable collision problem: public attitude

To be successfully introduced into the public domain, AVs must meet the expectations of the public (Adnan et al. 2018; Shariff, Bonnefon, and Rahwan 2017). Improving awareness of the benefits of AVs can vastly increase the probability that an individual will choose to purchase one over a traditional (SAE Level 0-1) vehicle (Haboucha, Ishaq, and Shiftan 2017). Without public acceptance, trust and willingness to buy and use, the true potential of AVs is unlikely to be realised. A prevalent study completed at the Massachusetts Institute of Technology (MIT) involved an online serious game which presented pairwise choices to participants, investigating what participants would want an AV to do in various situations (Awad et al. 2018). They collected 39.61 million decisions to pairwise choice scenarios from participants from 233 different countries. Analysis was performed to determine whether there is a preference for swerving or not swerving across scenarios with different variables including saving passengers or pedestrians, males or females, more lives or fewer lives, humans or pets, young or old, fit or less fit and social status. For pedestrians, it even considered whether they were crossing the road legally (Awad et al. 2018). This was a great example of how to involve the public to understand this idea of 'shared ethical principals'. However the results highlight a disparity in the range of what people consider to be ethical, with no overwhelming consensus.

Swerve versus don't swerve

The MIT Moral Machine Experiment found only a minor preference for staying true (Awad et al. 2018) as determined directly from the pairwise choices. Similar studies have shown this to be the case when the collision is due to be with one pedestrian, but have found little support for this being the case when the collision would be with ten pedestrians (Bonnefon, Shariff, and Rahwan 2016). Since the Moral Machine Experiment gives no information of this breakdown, we are unable to determine whether this contextual element trend occurred in their data also.

This swerve or no swerve decision is an important consideration as it takes us back to the issue of ambiguity over who is involved in the collision scenario. A clear preference for swerving to avoid ten pedestrians could indicate that AVs should be programmed to always swerve to avoid greater than a defined number of pedestrians. However, if this means colliding with a pedestrian on the pavement who is considered not to be involved in the scenario, then this would break one of the ethical guidelines (Luetge 2017).

Passengers versus pedestrians

When it comes to the decision of sacrificing a passenger or pedestrian, 76% participants in the study by Bonnefon, Shariff, and Rahwan (2016) expressed preference for sacrificing the passenger when ten pedestrians were involved. However, this reduced to 23% when one pedestrian would be collided with (Bonnefon, Shariff, and Rahwan 2016). The Moral Machine experiment found a minor preference for pedestrians over passengers (Awad et al. 2018). However Pickering, Podsiadly, and Burnham (2019) found that 92.4% of people chose to swerve into a barrier to save one pedestrian and 86.6% when they were told there were two people in the vehicle. The difference in this experiment was that participants were asked how the AV should be programmed, not what they would do, and were not told of the outcome of their decision i.e. death, injury etc. When it was a choice of one passenger or ten pedestrians, 96.3% stated preference for swerving into the barrier. However, these preferences changed when participants were aware of the consequences, participants instead choosing to minimise severity of injury or death regardless of the context (Pickering, Podsiadly, and Burnham 2019) – identifying a disparity in predicted acceptable outcomes compared to 'retrospective' acceptable outcomes.

It seems that people want their own vehicle to act to minimise the overall harm to human life, an ethical policy remnant of utilitarianism. This holds true only until the choice is between themselves and a single pedestrian where at that point, they want their AV to protect themselves. This choice is particularly true when the number of occupants in their vehicle increases (Awad et al. 2018; Bonnefon, Shariff, and Rahwan 2016; Morita and Managi 2020;

Dogan, Costantini, and Le Boennec 2020). Some even suggest that regulating AVs to be utilitarian to reduce overall casualties may in fact have the opposite effect if it causes reduced adoption of AVs (Smith 2019). Morita and Managi (2020) note that the credibility of AVs is very important, therefore it may be seen that as credibility of AV collision avoidance increases, the prominence of the desire to self-preserve may decrease. While this may be expected to some, the discussion around the Trolley Problem and Transplant Problem shows how even making choices to limit overall harm in the scenario is still morally questionable, and socially inconclusive. For the Trolley Problem it is seen most people think it is more moral to save the many and in the Transplant problem most think this is less moral (Gorr 1990). Hence, context can be very important in determining the moral action (Rhim, Lee, and Lee 2020). When consequences are given with absolute certainty some of the realism of the context of the scenario is lost (Rhim, Lee, and Lee 2020). Since most of the studies presented here are completed online, the 'emotional' element to the situation may also have been lost (Rhim, Lee, and Lee 2020) – the importance of which was highlighted when identifying the disparity between known outcomes and unknown outcomes (e.g., death or serious injury). Therefore, we must consider the reliability in the responses to these scenarios, and we are yet to fully understand what happens when context moves from hypothetical to real-world.

Considering this idea of swerve or no swerve, there seems to be a social trend in these experiments that total collateral damage is the prevalent factor on which the public want decisions to be made. Identifying the sum of this damage as part of this 'calculation' however is, as discussed, another problem in itself. One significant part of this is understanding the continuances of an action – as that forms part of the total 'cost' of any decision. For example, an avoided fatality has been shown to be valued 3.5 times higher than an avoided serious injury (Carlsson, Daruvala, and Jaldell 2010). To use this information however requires accurate injury assessment by the AV, which presents significant implementation challenges. Should the participant be given the choice between serious injury to the passenger or death of a single pedestrian, their preference may change. However, the decision to perform an action based on known consequences is different to that same decision with uncertainty added. It is important to consider the flaws in using certain outcomes to be able to effectively use the information from participant choices.

Effect of age

The previously discussed Moral Machine experiment also looked at the effect of age on preference, finding overall that people favour the young (Awad et al. 2018). Carlsson, Daruvala, and Jaldell (2010) found that avoiding the fatality of a 5-15 year old was seen to be the equivalent of 1.43 35-45 year olds and 3.31 65-75 year olds. This brings up another disparity between 'shared ethical principals' and the German ethics code as age would be a discriminatory factor which should not form part of the decision (Luetge 2017; Lin 2016). This is evidence of the potential disadvantage of using public viewpoint as a basis for the unavoidable collision problem. The highlighted preference for the young could be a consequence of natural bias, or procedural bias if age has a factor in willingness to take part in online surveys about AVs. It is important to understand the public viewpoint since the members of the public will be the consumers and users. However, the extent to which it is used in decisions for unavoidable collisions should be appropriately managed to avoid implementing unethical decisions as a result of bias. There is no evidence within the literature about the extent to which public preference should be prioritised, and instead only contradictions are highlighted in the guidelines around the need to understand shared ethical principles (although limiting the factors that these principals may be based on).

Other factors

Other factors introduced earlier in this discussion piece included considerations of humans or pets, gender, social status, fitness or health and legal or illegal crossing of the road (Awad et al. 2018). Firstly, and perhaps unsurprisingly, a strong preference for sacrificing pets over humans was found, showing alignment with the German Ethics Code (Awad et al. 2018; Luetge 2017). More worryingly however, a preference was shown for avoiding those of higher social status over avoiding those of lower social status (Awad et al. 2018). This once again is a preference that could be considered immoral by many outside of this experiment and calculating social status is an ethical dilemma in itself which should not be explored. It was shown that on average, the public have a preference in avoiding women over men, with 15% more choosing to save women, and for avoiding the fit over the unfit (Awad et al. 2018). These results amplify the point made in the subsection 'Effect of Age' of the contradiction between apparent public opinion and the Ethical guidelines presented in the section 'Current Ethical Guidelines'. If discriminatory bias is present in society, unconscious or otherwise, the introduction of AVs should be an opportunity to remove this bias, rather than implement it into machine code and cementing these socially damaging biases (e.g. the value of female life over male). A balance may need to be found between the use of public opinion and the desires of lawmakers and other stakeholders to ensure that solutions are representative. The final factor was that of legal versus illegal crossing of a road by a pedestrian. This has a further cultural context where in the UK, for example an unlawful crossing is more ambiguous, and pedestrians have the right to cross at any part of a road. This compared to the USA where laws are stricter on what is lawful and unlawful crossing. In this USA based study, participants showed greater preference in favour of the lawful (Awad et al. 2018). A similar preference was shown when other road users were acting unlawfully, participants preferring them to take a higher proportion of the negative consequences (Rhim, Lee, and Lee 2020). The use of this, if at all, is perhaps in the domain of legislative bodies who are in charge of defining the laws and the culture in which they sit, rather than those expected to adhere to them.

Different cultures may hold different ethical values and hence what might be considered moral in one culture may not in another (Awad et al. 2018). Rhim, Lee, and Lee (2020) showed differences in the reasoning within dilemma situations between Canadians and South Korean's for example. Awad et al. (2018) witnessed geographical dissimilarities in preference regarding age, social status, humans over pets, lawful versus unlawful road crossing and gender. Preference for the young and higher social status individuals is weaker in countries such as Japan and Indonesia and stronger in South American countries (Awad et al. 2018). An increasingly stronger preference for women and the fit is also witnessed in South American countries (Awad et al. 2018). However, through the use of ethical guidelines to narrow down the possible considerations in an unavoidable collision problem, it may be possible to find ethical principles that are shared by members of the public regardless of geographical location. A visual representation of the weighting applied to various characteristics in public opinion is shown in Figure 2.

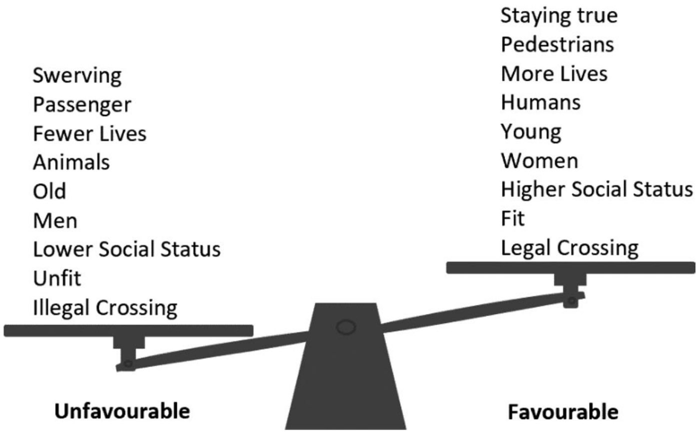

Figure 2. Visual representation of the weight applied to different characteristics in an unavoidable collision.

Agree to disagree?

The above section shows certain apparent preferences that are held within the public but, as noted in previous subsections, the extent to which they are used for decision making in an unavoidable collision should be scrutinised. The ethical guidelines provide a starting point, providing information on what an AV should not be able to do within an unavoidable collision (Luetge 2017; European Commission 2020). This leaves final considerations of whether to swerve or not swerve, pedestrian versus passenger, number of potential casualties, and the total 'damage' caused as the factors by which an AV can implement decisions based upon. It is important to best understand the public viewpoint on these factors alone. Whereas studies such as the Moral Machine Experiment change the characteristics of the actors involved with respect to age, gender etc., future studies may gain more benefit from omitting these factors. Although theoretically interesting, these factors have limited application and would only serve to divide society. While this reduces the complexity of the problem, it may not prevent public outcry through the collision with a child, for example. However, work on these factors will help prevent situations such as the collision with a pedestrian through the reduction of kinetic energy, when swerving into a wall could have avoided this. The public viewpoint on these issues will be essential to ensure the best balance between ethical actions and publicly accepted actions is met, if these do not coincide completely.

What we say versus what we do

We discussed previously the issue that people's responses in online surveys may not be representative of their true viewpoint as a result of them being too distant or abstracted from the scenario. To study this reliability issue a study by Bostyn, Sevenhant, and Roets (2018) involved the choice between giving an electric shock to one mouse to avoid shocking five mice, which was then conducted in the real-world to compare. It is important to note that no shocks were actually administered, although the participants did not know this

(Bostyn, Sevenhant, and Roets 2018). This is an adaptation of the 'Trolley Problem' whereby failure to act causes expected electric shocking of five mice but can be prevented by transferring it to one mouse. They found that the participants were twice as likely to refuse to switch the shocks in the hypothetical experiment as in the real-world experiment. A similar study by FeldmanHall et al. (2012) involved participants refusing the gift of money to stop an individual being given an electric shock (shocks were pre-recorded and a video was shown to participants), both in a hypothetical and real-world scenario (FeldmanHall et al. 2012). Participants were asked to make this choice 20 times, where £1 was at stake for each choice, totalling to £20. In the hypothetical scenario only 7% of participants kept a portion of the money, whereas 100% kept a portion of the money in the real case with a mean pay out of £12.52 and SD £5.43. In the real case, participants pressed a button at the end to randomly multiply the remaining money by between 1 and 10. In the hypothetical case, participants were told the remaining money would be multiplied by 10. The results show that self-benefit, in this case a relatively small amount money, can change the intention of participants to limit harm to others (FeldmanHall et al. 2012). These two studies demonstrate how much variance there can be between what we say and what we do and therefore question the reliability of self-report hypothetical solutions.

There have also been studies involving Virtual Reality (VR) which have tested this phenomenon (Francis et al. 2016; Patil et al. 2014). A study involving a footbridge dilemma where a participant can save five workers on a track by pushing an individual off a bridge was completed comparing choices in a hypothetical scenario with actions in VR (Francis et al. 2016). It was found that 10% of participants pushed the individual onto the tracks in the hypothetical scenario but this rose to 63.3% in the VR scenario. A different study using VR compared actions taken to text versions for 4 different types of moral dilemma (Patil et al. 2014). These included sacrificing one individual to save two and the same but saving five. The study showed that 76% chose to save the many in the text version and 95% in the VR version. The authors did note that the results varied based on whether the text choices were completed before or after the VR experience, pointing perhaps to an emotional response affecting the results (Patil et al. 2014). Hence, this backs our claim that adding emotion to the context of an unavoidable collision is important to get a better understanding of public viewpoint. The fact that what people say and what people do in moral dilemmas is not always equivalent shows that this must be taken into account when designing the study. While online studies are useful, it is important to acknowledge their limitations.

Ethical moral theory

Taking a step back from (online) experimental results and looking more theoretically at this problem, this paper has demonstrated throughout that it is not always obvious what the most moral action is in a given situation. As we saw with the Trolley Problem and Transplant Problem, changes in the context of the problem can cause objectively equivalent problems to have supposed alternative moral solutions (Gorr 1990). This demonstrates the complexity of morality. In this section we are going to explore different ethical moral theories. According to Bench-Capon, modern day ethical theories can be divided into three types: consequential ethics, deontological ethics, virtue ethics (Bench-Capon 2020).

Consequential ethics

The principle of consequential ethics is that the morality of an action is dependent upon the consequences of the action (Bench-Capon 2020; Dogan, Costantini, and Le Boennec 2020). In this moral theory, the reason for or process of deciding to perform an action has no effect on the morality of said action (Dogan, Costantini, and Le Boennec 2020). Utilitarianism is an example of consequential ethics whereby the action is considered moral as long as the consequence of the action is the minimising of harm to those involved parties (Dogan, Costantini, and Le Boennec 2020; Mordue, Yeung, and Wu 2020). The issue with this theory in the context of AVs is the lack of certainty over the consequences of an action (Bench-Capon 2020). We would need an AV to be able to predict the severity of outcomes in order to make a decision viewed as moral by consequentialism. Without certainty in this prediction, it is very difficult for the system to make moral decisions based on the consequences of actions (Bench-Capon 2020). However, advances in simulation and modelling (both of vehicles/crash structure and humans) may help us better understand such a prediction in the future (Perello-March et al. 2021).

In the case of our unavoidable collision described above, the solution proposed by consequential ethics may be to swerve into the side of the lorry, causing potential death to the occupant and damage to the lorry. However, this action could mean the vehicle collides with the oncoming vehicle anyway. So perhaps the solution proposed would be to collide head on with the oncoming vehicle, potentially killing both occupants. This demonstrates the difficulties of consequential ethics without certainty over the outcomes of decisions.

Deontological ethics

Deontological ethics is the theory that the morality of an action is determined by its conformity to a set of rules (Bench-Capon 2020; Dogan, Costantini, and Le Boennec 2020). Hence there must be a set of rules by which the morality of the action can be assessed. These rules will require foresight of all the possible circumstances and analysis of the most moral action in that circumstance in order to create rules by which an AV must follow (Bench-Capon 2020). Since an AV will by default follow these rules, it cannot disobey them based on context of the situation (Bench-Capon 2020). There is potential for disobeying of the rules to be programmed within the AV, and the morality of its decision to do so to be analysed in a court of law (Bench-Capon 2020). This is a useful theory for the implementation of a moral decision-making system, however, this does not help answer the question of what is the most moral action, in a defined context.

For example in a simplified case the rule could require the AV to always swerve off the road at the nearest point. In the context of our unavoidable collision above this would mean that the AV would swerve left and risk colliding with the pedestrian.

Virtue ethics

Virtue ethics suggests that good moral actions exhibit virtues and bad moral actions exhibit vices (Bench-Capon 2020; Mordue, Yeung, and Wu 2020). The moral action in a situation is assessed using values and levels of needs where lower level refers to more basic needs (Bench-Capon 2020). Bench-Capon states that to prefer one's higher level needs, over the lower level needs of another is selfish, the opposite is altruistic and to prefer another's higher level needs over one's lower level needs is sacrificial (Bench-Capon 2020). According to Bench-Capon, an ethical AV should be altruistic, may be sacrificial in some contexts and will not be selfish (Bench-Capon 2020). The idea here is that by defining levels of needs and perhaps even values, one can design ethical AVs which can assess morality in a specific context. This is an interesting framework on which to present future ethical decision-based research with the public. Understanding the needs of the involved parties in a particular context still, however, remains important (Bench-Capon 2020). The difference with virtue ethics is since the AV will have performed moral reasoning to make its decision, we can more accurately trace the reasoning for the behaviour (Bench-Capon 2020).

In the case of our unavoidable collision above, we can simplify the situation to 2 values: avoidance of harm and the avoidance of causing death. We also have the need of the occupant to preserve their life and to avoid harm to themselves. We could define the need to avoid personal harm and the value to not cause harm to others as the same level. Similarly for death to oneself or others, with this being a more basic need or value. Then swerving left would not be possible as this would likely cause death to the pedestrian but only harm to the occupant, which would be considered selfish. The other two options are less clear and hence would need prediction. The AV would have to demonstrate its decision-making process based on this definition of values and needs.

Other ethical policies

Other ethical policies mentioned in the related literature include Ethical Egoism (Mordue, Yeung, and Wu 2020). This principle is that actions are taken to maximise the benefit of one's self, the occupant in the context of an AV (Mordue, Yeung, and Wu 2020). This doesn't have to be at the detriment of others as it considers future hardship as well as short term gain, such as the effect of risking a pedestrian or the occupant (Mordue, Yeung, and Wu 2020). In the case of AV unavoidable collision's this may not cause immoral actions as those actions which are considered immoral are likely to cause greater psychological harm to the occupant, such as the lasting impact on an individual's mind from hitting a pedestrian. However, conversely, this could be viewed as a very selfish policy and hence innately immoral due to the policy being self-centred.

For the unavoidable collision problem described above, an AV operating Ethical Egoism would likely swerve left and risk hitting the pedestrian as this would reduce the harm to the occupant. However, psychological consequences may in fact mean that the AV should choose another option such as reducing kinetic energy as much as possible and colliding with the cause of the collision. It is this balance between physical and mental well-being that makes the decision as a result of Ethical Egoism ambiguous.

Discussion of moral theory

These ethical policies all have their advantages and disadvantages when assessed in the context of unavoidable collisions. Consequentialism has the benefit of judging based on the outcome of what would happen but has its disadvantages as these outcomes are not certain (Dogan, Costantini, and Le Boennec 2020; Bench-Capon 2020). Deontological ethics creates simplicity, as a set of rules are defined and the AV is considered moral provided it has followed these rules (Dogan, Costantini, and Le Boennec 2020; Bench-Capon 2020). However, this requires a completeness to the rules which again is difficult with numerous contexts and uncertain outcomes (Bench-Capon 2020). Virtue ethics implements decisions based on moral reasoning that could provide easier explanation in the event of a collision, but requires a set of socially universal moral values and needs on which to base this reasoning (Bench-Capon 2020; Mordue, Yeung, and Wu 2020). Ethical Egoism supports the public's apparent desire of self-preservation in their own AV but could be construed selfish or result in a greater number of fatalities (Mordue, Yeung, and Wu 2020).

Conclusion

The aim of this paper is to explore the work completed, both theoretical and 'experimental', towards identifying a common consensus on what an AV operating autonomously should do when faced with an unavoidable collision. The goal was not to present an ethical solution to the Trolley Problem which, by its nature, is unsolvable. Instead this paper examines the applicability of this thought experiment to future AV design, and critically analyses current work to identify a few gaps in knowledge for exploration. While the authors acknowledge that the aim is to avoid accidents entirely, it is unrealistically optimistic to assume there will be no collisions at all. Strategies such as reduction of kinetic energy may seem promising on face value but could be considered immoral in some contexts and hence do not offer an obvious solution. One of the first conclusions made in this regard is in evidencing a lack of experimental research through practical studies to be found in the literature. Throughout this paper we have referred to the MIT Moral Machine Experiment due to its size but also largely because there are few similar studies on which to compare. As with the MIT study, much of the research in this area is conducted by way of online surveys or equivalent, which provide a limited utility considering the lack of 'realism' which is a key factor for exploring morality. This paper identifies the need to understand the difference (if any) between what people may say they think is an ethical decision, versus what they would actually do when in such a scenario.

Regardless of ethical policy, we identify the need to involve the public to understand how they would want an AV to behave and what they consider to be ethical. Reviews of previous guidelines were identified as inconsiderate of true public opinion and were often contradictory of the public's views (albeit sometimes in a 'positive' way). Contradictions between experimental human factors studies and theoretical guidelines show that while useful, presented ethical guidelines do not provide the complete picture and will likely result in unacceptable outcomes in the public eye – damaging the acceptance of future autonomous technologies. In agreement with this conclusion, the ethical guidelines created by the European Commission talk of 'shared ethical principles' and the need for 'public deliberation' (European Commission 2020). This public deliberation is necessary to determine what these shared ethical principles are. However, this paper goes further than the European Commission through evidencing the need for participants to feel more attached to the decisions or choices they make regarding ethical dilemmas in unavoidable collisions. Techniques such as the use of driving simulators, for example the 3xD simulator at the University of Warwick, WMG, or virtual reality could be used to increase the reliability and validity of future research whilst ensuring safety. Only by ensuring that the responses to studies are representative of the views truly held can we understand fully the desires of the public for their future AVs.

Whilst usually a paper should draw conclusions based on its content, the content of this paper shows it is not appropriate for the authors to draw conclusions. Any conclusion the authors could have given that presents a solution to this problem would represent the opinion of the authors based on their own set of moral values – which as this paper has demonstrated may not be shared with others. Only with discussion and deliberation across all the involved stakeholders should a solution be found.

Acknowledgement

The authors acknowledge the contribution of Professor Stewart Birrell, Coventry University for his work supervising the research degree project that inspired this work.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Jonathan Robinson graduated with a BSc in Mathematics from the University of Warwick in 2018. He then completed an MSc in Smart, Connected and Autonomous Vehicles at WMG, University of Warwick in 2019, graduating in early 2020. Since 2019 he has been working as a Graduate Trainee Engineer at WMG, University of Warwick. He has been involved with work on the 3xD simulator at WMG, simulation of optical phased array LiDAR chips, programming-based tasks within projects and Human Factors related activities such as the creation of Human Machine Interfaces for mitigation of motion sickness.

Dr Joseph Smyth is a Research Fellow within the Intelligent Vehicles group at WMG in the National Automotive Innovation Centre. He received his BSc(hons) in Ergonomics and Human Factors from Loughborough University, an MA in Design Critical Practice from the University of London and an Engineering Doctorate (EngD) from the University of Warwick. Joseph's spent four years working alongside a British automotive OEM to combat motion sickness for future vehicles and motion sickness is still an ongoing research interest. His other research interests include driving simulation, trust and acceptance, driver state monitoring and human machine interface design for future mobility.

Dr Roger Woodman is an Assistant Professor at WMG, University of Warwick, where he leads the Human Factors research, within the Intelligent Vehicles group. He received his BSc in Computer Science from the University of Gloucestershire in 2006, MSc in Robotics from the University of the West of England in 2008, and PhD from the Bristol Robotics Laboratory in 2013. He has more than six years of industrial experience, working in leading manufacturing, defence, and software companies. He joined the University of Southampton in 2013 as a Research Fellow, specialising in biomedical imaging. In 2017, he joined WMG at the University of Warwick as a Research Fellow, investigating human factors of low-speed autonomous transport. Among his research interests, are shared mobility, provable AI, transport optimisation, micromobility, and last-mile logistics. He has several scientific papers published in the field of autonomous vehicles and robotics and is a Fellow of the HEA.

Dr Valentina Donzella is an Associate Professor at WMG, University of Warwick, and she has been awarded a Royal Academy of Engineering Industrial Fellowship (2020-2022). Valentina received her BSc (2003) and MSc (2005) in Electronics Engineering from University of Pisa and Sant'Anna School of Advanced Studies (Pisa, Italy), and her PhD (2010) in Innovative Technologies for Information,

Communication and Perception Engineering from Sant'Anna School of Advanced Studies. In 2009, she was a visiting graduate student at McMaster University (Hamilton, ON, Canada) in the Engineering Physics department. Before joining WMG, she was a MITACS and SiEPIC postdoctoral fellow at the University of British Columbia (Vancouver, BC, Canada), in the Silicon Photonics group. She is first author and co-author of several journal papers on top tier optics journals. Her research interests are: LiDAR, Intelligent Vehicles, integrated optical sensors, sensor fusion, and silicon photonics. Dr Donzella is Full College member of EPSRC and Senior Fellow of Higher Education Academy.

ORCID

Joseph Smyth https://orcid.org/0000-0001-7036-1045

References

Adnan, N., S. Md Nordin, M. Ariff bin Bahruddin, and M. Ali. 2018. "How Trust Can Drive Forward the User Acceptance to the Technology? In-Vehicle Technology for Autonomous Vehicle." *Transportation Research Part A: Policy and Practice* 118: 819–836. doi:10.1016/j.tra.2018.10.019.

Anderson, J. M., K. Nidhi, K. D. Stanley, O. A. Oluwatola, C. Samaras, and P. Sorensen. 2014. *Autonomous Vehicle Technology: A Guide for Policymakers*. Santa Monica: RAND Corporation, The.

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J-F. Bonnefon, and I. Rahwan. 2018. "The Moral Machine Experiment." *Nature* 563 (7729): 59–64. doi:10.1038/s41586-018-0637-6.

BBC (British Broadcasting Company). 2020. "Uber's Self-Driving Operator Charged over Fatal Crash." https://www.bbc.co.uk/news/technology-54175359.

Bench-Capon, T. J. M. 2020. "Ethical Approaches and Autonomous Systems." *Artificial Intelligence* 281: 103239. doi:10.1016/j.artint.2020.103239.

Bonnefon, J-F, A. Shariff, and I. Rahwan. 2016. "The Social Dilemma of Autonomous Vehicles." *Science (New York, N.Y.)* 352 (6293): 1573–1576. doi:10.1126/science.aaf2654.

Bostyn, D. H., S. Sevenhant, and A. Roets. 2018. "Of Mice, Men, and Trolleys: Hypothetical Judgment versus Real-Life Behavior in Trolley-Style Moral Dilemmas." *Psychological Science* 29 (7): 1084–1093. doi:10.1177/0956797617752640.

Burton, S., I. Habli, T. Lawton, J. McDermid, P. Morgan, and Z. Porter. 2020. "Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective." *Artificial Intelligence* 279: 103201. doi:10.1016/j.artint.2019.103201.

Carlsson, F., D. Daruvala, and H. Jaldell. 2010. "Preferences for Lives, Injuries, and Age: A Stated Preference Survey." *Accident; Analysis and Prevention* 42 (6): 1814–1821. doi:10.1016/j.aap.2010.05.002.

Conger, K. 2020. "Driver Charged in Uber's Fatal 2018 Autonomous Car Crash." *The New York Times*. https://www.nytimes.com/2020/09/15/technology/uber-autonomous-crash-driver-charged.html

Dimitrakopoulos, G., L. Uden, and I. Varlamis. 2020. "User Acceptance and Ethics of ITS." In *The Future of Intelligent Transport Systems, edited by G. Dimitrakopoulos, L. Uden, and I. Varlamis*, 85–91. Elsevier. doi:10.1016/b978-0-12-818281-9.00007-3..

Dogan, E., F. Costantini, and R. Le Boennec. 2020. "Ethical Issues concerning Automated Vehicles and Their Implications for Transport." *Advances in Transport Policy and Planning* 5: 215–233. doi:10.1016/bs.atpp.2020.05.003..

European Commission. 2020. "Ethics of Connected and Automated Vehicles: Recommendations on Road Safety, Privacy, Fairness, Explainability and Responsibility." doi:10.2777/966923..

Fagnant, D. J., and K. Kockelman. 2015. "Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations." *Transportation Research Part A: Policy and Practice* 77: 167–181. doi:10.1016/j.tra.2015.04.003.

FeldmanHall, O., D. Mobbs, D. Evans, L. Hiscox, L. Navrady, and T. Dalgleish. 2012. "What We Say and What We Do: The Relationship between Real and Hypothetical Moral Choices." *Cognition* 123 (3): 434–441. doi:10.1016/j.cognition.2012.02.001.

Flemisch, F., M. Heesen, T. Hesse, J. Kelsch, A. Schieben, and J. Beller. 2012. "Towards a Dynamic Balance between Humans and Automation: Authority, Ability, Responsibility and Control in Shared and Cooperative Control Situations." *Cognition, Technology & Work* 14 (1): 3–18. doi:10.1007/s10111-011-0191-6.

Francis, K. B., C. Howard, I. S. Howard, M. Gummerum, G. Ganis, G. Anderson, and S. Terbeck. 2016. "Virtual Morality: Transitioning from Moral Judgment to Moral Action?" *PLoS ONE* 11 (10): e0164374–23. doi:10.1371/journal.pone.0164374.

Gorr, M. 1990. "Thomson and the Trolley Problem." *Philosophical Studies* 59 (1): 91–100. doi:10.1007/BF00368393.

Haboucha, C. J., R. Ishaq, and Y. Shiftan. 2017. "User Preferences regarding Autonomous Vehicles." *Transportation Research Part C: Emerging Technologies* 78: 37–49. doi:10.1016/j.trc.2017.01.010.

Honda European Media Newsroom. 2021. "Honda Launches Next Generation Honda Sensing Elite Safety System with Level 3 Automated Driving Features." https://hondanews.eu/eu/et/corporate/media/pressreleases/329456/honda-launches-next-generation-honda-sensing-elite-safety-system-with-level-3-automated-driving-feat

Lee, D. 2019. "Uber Self-Driving Crash 'Mostly Caused by Human Error.'" *BBC*. https://www.bbc.co.uk/news/technology-50484172

Leslie, D. 2019. "Understanding Artificial Intelligence Ethics and Safety Systems in the Public Sector." *The Alan Turing Institute*.

Levin, S., and J. C. Wong. 2018. "Self-Driving Uber Kills Arizona Woman in First Fatal Crash Involving Pedestrian." *The Guardian*. https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe

Lin, P. 2016. "Why Ethics Matters for Autonomous Cars." In *Autonomous Driving: Technical, Legal and Social Aspects*, edited by M. Maurer, J. Gerdes, B. Lenz, and H. Winner, 69–85. Berlin, Heidelberg: Springer. doi:10.1007/978-3-662-48847-8.

Luetge, C. 2017. "The German Ethics Code for Automated and Connected Driving." *Philosophy & Technology* 30 (4): 547–558. doi:10.1007/s13347-017-0284-0.

Martínez-Díaz, M., and F. Soriguera. 2018. "Autonomous Vehicles: Theoretical and Practical Challenges." *Transportation Research Procedia* 33: 275–282. doi:10.1016/j.trpro.2018.10.103.

McCausland, P. 2019. "Self-Driving Uber Car That Hit and Killed Woman Did Not Recognize That Pedestrains Jaywalk." *NBC News*.https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

Mordue, G., A. Yeung, and F. Wu. 2020. "The Looming Challenges of Regulating High Level Autonomous Vehicles." *Transportation Research Part A: Policy and Practice* 132: 174–187. doi:10.1016/j.tra.2019.11.007.

Morita, T., and S. Managi. 2020. "Autonomous Vehicles: Willingness to Pay and the Social Dilemma." *Transportation Research Part C: Emerging Technologies* 119: 102748. doi:10.1016/j.trc.2020.102748.

Patil, I., C. Cogoni, N. Zangrando, L. Chittaro, and G. Silani. 2014. "Affective Basis of Judgment-Behavior Discrepancy in Virtual Experiences of Moral Dilemmas." *Social Neuroscience* 9 (1): 94–107. doi:10.1080/17470919.2013.870091.

Payre, W., S. Birrell, and A. M. Parkes. 2021. "Although Autonomous Cars Are Not yet Manufactured, Their Acceptance Already Is." *Theoretical Issues in Ergonomics Science* 22 (5): 567–514. doi:10.1080/1463922X.2020.1836284.

Perello-March, J. R., C. G. Burns, R. Woodman, M. T. Elliott, and S. A. Birrell. 2021. "Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios." *IEEE Transactions on Intelligent Transportation Systems* 1–11. doi:10.1109/TITS.2021.3050518.

Pickering, J. E., M. Podsiadly, and K. J. Burnham. 2019. "A Model-to-Decision Approach for the Autonomous Vehicle (AV) Ethical Dilemma: AV Collision with a Barrier/Pedestrian(S)." *IFAC-PapersOnLine* 52 (8): 257–286. doi:10.1016/j.ifacol.2019.08.080.

Rhim, J., G-b. Lee, and J-H. Lee. 2020. "Human Moral Reasoning Types in Autonomous Vehicle Moral Dilemma: A Cross-Cultural Comparison of Korea and Canada." *Computers in Human Behavior* 102: 39–56. doi:10.1016/j.chb.2019.08.010.

Robles Carrillo, M. 2020. "Artificial Intelligence: From Ethics to Law." *Telecommunications Policy* 44 (6): 101937. doi:10.1016/j.telpol.2020.101937.

SAE (Society of Automotive Engineers). 2021. "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles." https://www.sae.org/standards/content/j3016_202104/

Schiff, D., J. Biddle, J. Borenstein, and K. Laas. 2020. "What's Next for AI Ethics, Policy, and Governance? A Global Overview." AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 153–58. New York, NY, USA. doi:10.1145/3375627.3375804.

Shariff, A., J-F. Bonnefon, and I. Rahwan. 2017. "Psychological Roadblocks to the Adoption of Self-Driving Vehicles." *Nature Human Behaviour* 1 (10): 694–696. doi:10.1038/s41562-017-0202-6.

Smith, B. 2019. "Personality Facets and Ethics Positions as Directives for Self-Driving Vehicles." *Technology in Society* 57: 115–124. doi:10.1016/j.techsoc.2018.12.006.

Smyth, J., P. Jennings, and S. Birrell. 2020. "Are You Sitting Comfortably? How Current Self-Driving Car Concepts Overlook Motion Sickness, and the Impact It Has on Comfort an Productivity." In *Advances in Human Factors of Transportation. AHFE 2019. Advances in Intelligent Systems and Computing*, edited by N. Stanton, Vol. 964, 387–399. Springer, Cham. doi:10.1007/978-3-030-20503-4.