

The European Commission report on ethics of connected and automated vehicles and the future of ethics of transportation

Santoni de Sio, Filippo

DOI

[10.1007/s10676-021-09609-8](https://doi.org/10.1007/s10676-021-09609-8)

Publication date

2021

Document Version

Final published version

Published in

Ethics and Information Technology

Citation (APA)

Santoni de Sio, F. (2021). The European Commission report on ethics of connected and automated vehicles and the future of ethics of transportation. *Ethics and Information Technology*, 23(4), 713-726.
<https://doi.org/10.1007/s10676-021-09609-8>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



The European Commission report on ethics of connected and automated vehicles and the future of ethics of transportation

Filippo Santoni de Sio¹

Accepted: 19 July 2021
© The Author(s) 2021

Abstract

The paper has two goals. The first is presenting the main results of the recent report *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility* written by the Horizon 2020 European Commission Expert Group to advise on specific ethical issues raised by driverless mobility, of which the author of this paper has been member and rapporteur. The second is presenting some broader ethical and philosophical implications of these recommendations, and using these to contribute to the establishment of Ethics of Transportation as an independent branch of applied ethics. The recent debate on the ethics of Connected and Automated Vehicles (CAVs) presents a paradox and an opportunity. The paradox is the presence of a flourishing debate on the ethics of one very specific transportation technology without ethics of transportation being in itself a well-established academic discipline. The opportunity is that now that a spotlight has been switched on the ethical dimensions of CAVs it may be easier to establish a broader debate on ethics of transportation. While the 20 recommendations of the EU report are grouped in three macro-areas: road safety, data ethics, and responsibility, in this paper they will be grouped according to eight philosophical themes: Responsible Innovation, road justice, road safety, freedom, human control, privacy, data fairness, responsibility. These are proposed as the first topics for a new ethics of transportation.

Keywords European Commission Report on ethics of CAVs · Ethics of self-driving cars · Ethics of transportation · Responsible innovation in self-driving cars

The recent debate on the ethics of Connected and Automated Vehicles (henceforth CAVs)¹ presents a paradox and an opportunity. The paradox is that there is now a flourishing debate on the ethics of one very specific, only partially existing, transportation technology—CAVs—without ethics of transportation in general being in itself a well-established, independent academic discipline. The opportunity is that now that a spotlight has been switched on the ethical dimensions of CAVs it may be easier to establish a broader and more systematic debate on ethics of transportation. This paper has two goals. The first is descriptive, limited and backward-looking: presenting the main results of the recent report *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility* written by the Horizon 2020 Commission

Expert Group to advise on specific ethical issues raised by driverless mobility, of which the author of this paper has been member and rapporteur (Bonnefon et al., 2020) (henceforth “EC Report on Ethics of CAVs” or “The EC Report”).² The second is broader, open-ended, and forward-looking: presenting some ethical and philosophical issues behind these recommendations, and using these to encourage the

✉ Filippo Santoni de Sio
f.santonidesio@tudelft.nl

¹ Section Ethics/Philosophy of Technology, Delft University of Technology, Delft, The Netherlands

¹ Aka Self-driving cars, automated driving systems, autonomous vehicles etc. The paper will follow the EU report in the use the term: Connected and Automated Vehicles (CAVs). In the report CAVs are defined as “Vehicles that are both connected and automated and display one of the five levels of automation according to SAE International’s standard J3016, combined with the capacity to receive and/or send wireless information to improve the vehicle’s automated capabilities and enhance its contextual awareness.” (Bonnefon et al., 2020, p. 12).

² The presentation and discussion of the content of the EC Report on Ethics of CAVs contained in this paper is necessarily partial and reflects the view of the author of this paper and not necessarily that of the entire Expert Group. The reader is referred to the report for the original presentation of the recommendations and their discussion.

establishment of Ethics of Transportation as an independent academic discipline.

The paper is structured as follows. The next “[The European Commission Report on Ethics of CAVs](#)” section briefly summarises the background and main content of the EC Report on Ethics of CAVs. “[Ethics of CAVs and the future of ethics of transportation](#)” section proposes a brief history of ethics of CAVs and makes a plea for the establishment of ethics of transportation as an independent domain of applied ethics. Each of the eight successive sections discusses a set of recommendations of the EC Report on Ethics of CAVs. While in the EC Report the 20 recommendations are grouped in three macro-areas: road safety, data ethics, responsibility, in this paper they will be grouped according to eight philosophical themes: Responsible Innovation, road justice, road safety, freedom, meaningful human control, privacy, data fairness, responsibility. These are proposed as the first topics for a new ethics of transportation. For each of these sets, the broader ethical and philosophical background is presented, the broader relevance for ethics of transportation is discussed, and some suggestions for (new) research questions in ethics of transportation are given. The final section summarises and looks ahead to future research.

The European Commission report on ethics of CAVs

In September 2017 the Ethics Task Force, a Member State initiative that was set up after the second High Level Meeting of EU Transport Ministers, the European Commission and Industry on Connected and Automated Driving in Frankfurt, emphasized “The need to discuss ethical issues raised by Connected and Automated Vehicles (CAVs) at European level” (Di Fabio et al., 2018). In a 2018 communication, the European Commission announced the creation of a Commission Expert Group to advise on specific ethical issues raised by driverless mobility.” (On the Road to Automated Mobility: An EU Strategy for Mobility of the Future, 2018).

The group was established as an Independent Expert Group³ in 2019 under the supervision of the EU's Directorate-General for Research and Innovation. It delivered its report in June 2020, after six formal meetings, including a stakeholder workshop. The group was composed by Jean-Francois Bonnefon (chairman), social psychologist famous for his work on the ethical dilemmas of self-driving cars, Filippo Santoni de Sio (rapporteur), academic philosopher

specialised in ethics of technology, and twelve more experts in ethics of CAVs with different disciplinary profiles: three more moral philosophers (ethicists): David černý, John Danaher, Stavroula Tsinorema, two lawyers: Nathalie Devillier and Sandra Wachter, one data scientist: Veronika Johansson, one traffic psychologist: Marieke Martens, one social scientist: Karolina Zawieska, and four transport engineers: Tatiana Kovacikova, Miloš Mladenovič, Paula Palade, Nick Reed.

The report contains the following twenty recommendations. Each recommendation includes specific general tasks for researchers, manufacturers and policy-makers as well as some discussion on the theoretical and practical reasons behind the recommendation.

Road safety

1. Ensure that CAVs reduce physical harm to persons.
2. Prevent unsafe use by inherently safe design.
3. Define clear standards for responsible open road testing.
4. Consider revision of traffic rules to promote safety of CAVs and investigate exceptions to non-compliance with existing rules by CAVs.
5. Redress inequalities in vulnerability among road users
6. Manage dilemmas by principles of risk distribution and shared ethical principles.

Data and algorithm ethics: privacy, fairness, and explainability

7. Safeguard informational privacy and informed consent.
8. Enable user choice, seek informed consent options and develop related best practice industry standards.
9. Develop measures to foster protection of individuals at group level.
10. Develop transparency strategies to inform users and pedestrians about data collection and associated rights.
11. Prevent discriminatory differential service provision.
12. Audit CAV algorithms.
13. Identify and protect CAV relevant high-value datasets as public and open infrastructural resources.
14. Reduce opacity in algorithmic decisions.
15. Promote data, algorithmic, AI literacy and public participation.

Responsibility

16. Identify the obligations of different agents involved in CAVs.
17. Promote a culture of responsibility with respect to the obligations associated with CAVs.
18. Ensure accountability for the behaviour of CAVs (duty to explain).

³ This means the experts were speaking in their own individual capacity and not on behalf of the Commission, or the organizations they are part of.

19. Promote a fair system for the attribution of moral and legal culpability for the behaviour of CAVs.
20. Create fair and effective mechanisms for granting compensation to victims of crashes or other accidents involving CAVs.

Ethics of CAVs and the future of ethics of transportation

Ethics of CAVs has also an academic history. In 2014, Jason Millar, Patrick Lin and other scholars in ethics of technology started publishing a series of popular and academic articles on the so-called ethical dilemmas of self-driving cars (Lin, 2014, 2015; Millar, 2014a, 2014b). These articles presented different versions of thought experiments in which vehicles equipped with automated driving systems face unavoidable crashes, where different outcomes are open, typically crashing into a wall thereby injuring or killing the car occupants or swerving and hitting different persons or animals. The questions are posed: who or what should the vehicle hit and who should take the decision. The papers were a variation on the famous trolley problems in moral philosophy (Thomson, 1985), and were meant to raise a debate on some ethical dilemmas that may emerge with the growing use of automated systems in society. At a more general level, they were also meant to highlight the older issue of non-neutrality of technology, that is the fact that technological design always (implicitly) embeds some normative principles that will eventually make them encourage or determine some uses and discourage or prevent others, benefit some individuals or groups and penalize others (Winner, 1980). In those same years, Jean Francois Bonnefon and some social psychologists from MIT launched the moral machine experiment (Bonnefon et al., 2016), in which subjects were asked to express online their moral judgements about decisions in scenarios similar to those imagined by Millar, Lin and others. The massive participation to the experiment and its broad media coverage gave an additional push to the publication of many more ethical articles on the topic (Nyholm, 2018a, 2018b).

However, a new wave of papers took a more critical attitude toward the “trolleyology” of self-driving cars (Nyholm & Smids, 2016) and turned to more general, and arguably more urgent, ethical issues with their introduction. These included risk and safety (Goodall, 2016), distributive justice (Mladenović, 2017), rights and inequalities (Liu, 2017), human control (Mecacci & Santoni de Sio, 2020), responsibility (Hevelke & Nida-Rümelin, 2014) the political dimension of vehicle automation (Himmelreich, 2019; Jafari-Naimi, 2018; Stilgoe, 2017). The first fatalities caused by vehicles equipped with automated driving systems (Bellon, 2018; Yadron & Tynan, 2016) offered a tragic confirmation

that driving automation was a source not only of interesting thought experiments about futuristic scenarios with fully autonomous vehicles, but also of very urgent ethical and practical problems with the introduction of less-than-fully autonomous vehicles on the road.

The great attention to ethics of CAVs emerged in the past few years presents a paradox and an opportunity. The paradox is that there is now a flourishing debate on the ethics of one very specific, only partially existing, transportation technology without ethics of transportation in general being in itself a well-established discipline. It is a bit as if we had an ethical debate about surgical robots without a well-established medical ethics, or an ethical debate on autonomous weapon systems, without an ethics of armed conflicts.⁴ To be fair, there do exist some books and many articles discussing important issues in ethics of transportation, especially in relation to justice or fairness (Martens, 2017; Pereira et al., 2017; Van Wee, 2011), and road safety in general (Hansson, 2014; Nihlén Fahlquist, 2009b; Ori, 2014, 2020; Rajan, 2007; Smids, 2018). And some scholars have already made the general suggestion to take ethics of transportation (Hosmer, 1996; Richardson, 1995) or road safety (Nihlén Fahlquist, 2009a; Ori, 2020) more seriously. But this has not yet result in the creation of ethics of transportation as a full-blown, independent, domain of applied ethics. At the time of writing, to the best of the author’s knowledge, there is no academic journal for ethics of transportation, and no academic courses on this topic, other than his own. There are certainly good historical reasons for this, the most important probably being that unlike for instance the medical and the military, there is not such a thing as a transportation profession, meant as a well-defined body of experts with shared education curricula, organisations, representatives etc.⁵

Be that as it may, this explanation does not amount to a justification for this gap. Far from it. Just to make a couple of examples: while rich and industrialized countries debate on how the introduction of the new CAVs technologies might (positively) affect the safety, comfort, sustainability, accessibility of road traffic, many developing countries still lag behind in the introduction of existing technical and social infrastructures that may dramatically reduce the number of (fatal) accidents, especially among the most vulnerable road users. The latter issue seems to raise issues of justice that are

⁴ In fact, ethics of CAVs may be seen also as part of ethics of AI and robotics. In this sense, ethics of CAVs is at the intersection between ethics of robotics and ethics of transportation.

⁵ Husak (2004) and Ori (2020) attribute this to the reluctance of philosophers to address new themes, especially those with a strong empirical or component, but this hypothesis doesn’t account for the recent emergence of other new branches of applied ethics with similar features, such as environmental ethics, ethics of technology or ethics of AI.

certainly not less relevant than those raised by vehicle automation (Nihlén Fahlquist, 2009b). Or, even within industrialised countries, existing technologies that could dramatically improve road safety, e.g. speed limiting technologies, are not compulsory, mainly due to concerns for individual freedom (Smids, 2018). To what extent this preference of individual freedom over public safety is justified, is another example of a very urgent and important ethical and philosophical question, which deserves more attention. A broader list of these issues will be presented and discussed below.

In this sense, the recent debate on the ethics of CAVs presents also an opportunity. Now that a spotlight has been switched on the many ethical dimensions of one specific transportation technology—CAVs—it may be a good moment to start a broader and more systematic philosophical reflection on such issues in ethics of transportation more generally.

New domains of applied ethics have often been born due to contingent, catastrophic historical circumstances. The view of the thousands of (civilian) victims of the battle of Solferino in 1859 pushed Henri Dunant to campaign for the establishment of what are today known as the Geneva Conventions, which in turn gave a strong impulse to the establishment of just war theory and military ethics as new domains of applied ethics. Similarly, medical and research ethics were established as academic disciplines in Europe after WWII, following the horrors of the inhuman scientific and medical experiments done by the Nazi regime. At least in terms of fatalities numbers, road safety and justice are certainly a humanitarian emergency comparable to a war. The historian Peter Norton noted that more Americans died in road accidents in the 4 years after World War I than were killed in combat in France (Norton, 2008). And, more recently, the legal philosopher Douglas Husak noted that in September 2001 more people died in the US due to road accidents than due to the attacks to the World Trade Center (Husak, 2004). Still, we don't have an ethics of transportation. Academic history has its own ways. Maybe one far day, people will be look back at the time of the ethical debates on automated driving, and will realise that, whatever the destiny of automated driving and its ethics will have been, these will have had the merit of contributing to the establishment of something bigger and more important: ethics of transportation as a new domain of applied ethics.

Notwithstanding its big ambition of contributing to the establishment of ethics of transportation as a new research domain, the paper has a necessarily limited scope. Transportation, in its broader sense of mobility, raises far-ranging ethical, legal, and political issues (Sager, 2016). The migrant crises and the debate over its regulation clearly shows that (not) being allowed or supported to safely move across countries, continents and seas is for many a matter of life-and-death. In this sense, to what extent people should have a

right to freely and safely move across the world is an issue in *ethics of mobility*. The ethics of tourism—how people should behave while travelling in foreign countries—might be another example of a topic in ethics of mobility, broadly construed. However, this paper will take a narrower perspective, and will focus only on *ethics of transportation*. Ethics of transportation is here defined as the reflection on the values, principles, norms, concepts that should guide the design, production, regulation, and use of different *transportation systems and infrastructures* in different contexts. In this sense, it is interesting to note that while ethics of transportation covers a narrower area than ethics of mobility, it still covers a broader area than professional ethics such as military or medical ethics. Ethics of transportation is not only an ethics for professionals but also for users of transportation systems and infrastructures, and for all persons more generally.

Moreover, ethics of transportation covers a broad and diverse range of domains, from road traffic, railway, aviation, marine, and even space technologies. However, this paper, by taking the lead from the EC Report on Ethics of CAVs, will only focus on the ethics of road traffic. Future research will have to map and start discussing ethics of transportation in the other domains.

Responsible innovation, CAVs, and ethics of transportation

The EC Report's goal, as state in its introduction is “supporting stakeholders in the systematic inclusion of ethical considerations in the development and regulation of CAVs.” (Bonnefon et al., 2020) p. 15). The report endorses and relies upon the Responsible Innovation approach to technological development, according to which an inclusive and responsive dialogue between stakeholders is key to the inclusion of ethical considerations in technological development (Stilgoe et al., 2013). A similar endorsement of the Responsible Innovation approach could be found in some of the major references of the report: the European Group on Ethics in Science and New Technology statement (European Group on Ethics in Science and New Technologies, 2018), the AIHLEG Guidelines for Trustworthy AI (European Commission, 2019), and the Dutch White Paper on Ethics and Self-driving Cars (Santoni de Sio, 2016). The willingness of the EC Expert Group to concretely realise this ideal is confirmed by them organising one stakeholder meeting already during preparation of the report, where “researchers, policymakers, associations and industry experts ... received a draft report with recommendations upon which they could propose revisions” (Bonnefon et al., 2020, p. 15). The report also warns that its recommendations “can and should be further discussed in future stakeholder meetings, on the basis of

further data and experience in the development and deployment of CAVs” (p. 19).

According to the authors, this ethics report is needed in two ways. First, “legislation alone may be insufficient to ensure that the development, deployment and use of CAVs is aligned with ethical principles and norms” (p. 16). Also in relation to CAVs, the ideal of Responsible Innovation can be realised only with the active contributions of all stakeholders: policy-makers as well as manufacturers, deployers, and users. Second, ethical issues with new technologies such AI have been already discussed in various policy and regulatory reports, such as the GDPR for data protection, the AIHLEG report for AI in general, the dedicated Expert group for Liability with AI (*Liability for Artificial Intelligence and Other Emerging Digital Technologies*, 2019). Also, some ethical issues with CAVs had been addressed by some other reports, such as the German Ethics Commission (Di Fabio et al., 2018), and the Ethics Task Force (e.g. moral dilemmas with crash avoidance, distribution of responsibility in complex networks) (Task Force on Ethical Aspects of Connected and Automated Driving, 2018). However, while relying on previous reports, the EC Report tries to answer some open questions, as well as raising new issues.

In the same months in which the EC Report was published, one important representative of the Responsible Innovation approach, Jack Stilgoe, published a short book, *Who’s driving Innovation*, in which among other things he endorsed the program of applying the RI approach to self-driving cars (Stilgoe, 2020). Stilgoe has always been very vocal in his opposition to “solutionist” narratives according to which (new) technologies may solve societal challenges on their own. In fact, no matter how higher their level of intelligence and automation, technologies will never be “autonomous” as they will always reflect and express specific goals and values of their human manufacturers, developers, users (Mindell, 2015; Stilgoe, 2017). These technologies can be therefore societally beneficial only to the extent to which they will be designed to support and respond to the relevant societal needs and values. In line with this anti-solutionist approach, as well as with idea of Design for Values, as expressed for instance by (Van den Hoven et al., 2012), the EC report explicitly frames ethical principles and values not as much as constraints to technological innovation, but rather as drivers for a technology that is able to explicitly address thorny and urgent societal challenges and to respond to deep human needs.

One recurring objection to the idea of Design for Values is that people tend to (legitimately) disagree on normative matters, and it is therefore unclear which values and principles should be embedded in the technological developments (e.g. (Manders-Huits, 2011)). In this respect, following the EGE’s statement on Artificial Intelligence and the AIHLEG’s Guidelines for Trustworthy AI, the EC report

claims that, as a matter of fact, a European agenda for Responsible Innovation for CAVs does have a clear, uncontroversial, normative starting point. This is represented by the “fundamental ethical and legal principles that have been critically and reflexively adopted by society, and that are laid down in the EU Treaties and in the EU Charter of Fundamental Rights” (Bonnefon et al., 2020, p. 21). Accordingly, before moving to the discussion of the twenty recommendations, the report presents the major “guiding ethical principles” of the Report and their relevance for CAVs. These are: Non-maleficence (CAVs should not increase the risk of harm for road users), Beneficence (CAV should contribute positively to the welfare of people and the environment), Dignity (CAVs should not infringe human rights or sacrifice them in the name of other societal goals), Personal Autonomy (CAVs should respect and promote people’s free agency and diverse ways of life), Justice (CAVs should provide equality of access to mobility for all), Responsibility (human persons should remain ultimately responsible for the behaviour of CAVs), Solidarity (CAVs should promote social collaboration toward shared goals for instance by sharing data about fatalities and injuries to improve road safety). These principles cannot be translated into practice with a “mechanical top-down procedure”, and they will have to be “specified, discussed and redefined in-context” (p. 23). That is why the introductory part of the report closes where it has started, that is with the principle of Inclusive Deliberation: CAV systems should be supportive of- and resulting from inclusive deliberation processes involving relevant stakeholders and the wider public.

While presented with the intention to be applied to CAVs, the methodology and principles presented in the report’s introductory parts may well be taken as a manifesto for the creation of a European ethics of transportation technologies more generally.⁶ Transportation systems are complex “socio-technical systems” whose functioning dramatically affects the life of most people in society. Designing transportation systems—automated or not—is therefore a big ethical and political enterprise. How to ensure that these systems don’t endanger but promote basic shared ethical and legal principles is an enterprise that has so far received less attention than it should.

⁶ From the point of view of a broader, global perspective, the unique reference to European values may not be sufficient. Integrating different cultural and political perspectives will be an additional task of a future ethics of transportation.

Justice in transportation: CAVs and beyond

Two important recommendations from the EC Report deal with issues of justice and fairness in road safety. Recommendation 5 (“Redress inequalities in vulnerability among road users”) states that “in order to address current and historic inequalities of road safety, CAVs may be required to behave differently around some categories of road users, e.g. pedestrians or cyclists, so as to grant them the same level of protection as other road users. CAVs should, among other things, adapt their behaviour around vulnerable road users instead of expecting these users to adapt to the (new) dangers of the road.” To be clear, the idea is not that vulnerable users should receive higher levels of safety, but rather that given their vulnerability, they should be treated differently, in order to enjoy the same level of safety of other road users. As explained in the report, the basic design concept behind this is a change of focus “from vulnerable users needing to adapt to the dangers of the road to CAVs needing to adapt to vulnerable road users” (p. 31). The authors of the report are aware of the technical and political challenges behind this idea and presents this as a topic for further research and discussion.

Recommendation 6 connects justice with the debate over moral dilemmas in unavoidable crash situations (see “[The European Commission Report on Ethics of CAVs](#)” section above). It states that “While it may be impossible to regulate the exact behaviour of CAVs in unavoidable crash situations, CAV behaviour may be considered ethical in these situations provided it emerges organically from a continuous statistical distribution of risk by the CAV in the pursuit of improved road safety and equality between categories of road users.”. This is an interesting move. In line with the position of the Ethics Task force and the Dutch White Paper, the authors of the report explicitly endorse the claim that “moral dilemmas in crash avoidance are not the only, nor even the most urgent, ethical and societal issue raised by CAV safety” (17). Therefore, the report proposes to let “moral dilemmas” fall under the general umbrella of issues of justice in road safety, and just recommends ensuring that the behaviour of CAVs in morally dilemmatic scenarios does not violate any of the ethical and legal principles proposed in the report. These are: justice as discussed in recommendation 5 above (equal safety for vulnerable users), dignity (no person or group should be explicitly or implicitly discriminated by design in the collision management), public deliberation (people’s concerns about the idea of CAVs taking life-and-death decision should be carefully and explicitly addressed in the public debate), solidarity (when compatible principles of privacy, data about the outcomes of dilemmas should be shared by safety agencies).

Among the topics in ethics of transportation, justice and fairness are probably the ones who have so far received the most attention (Pereira et al., 2017) with the few existing books in this field dealing with this topic (Jeekel, 2018; Martens, 2017; Van Wee, 2011). These authors have converged on the idea of going beyond traditional Cost–Benefit Analysis as the only or main policy tool in transportation policies, due to its inability to sufficiently accommodate important concerns with fairness and justice. Same for purely market-driven libertarian approaches. However, important questions still need to be answered: should road safety be considered, in a deontological fashion, a basic right, according to the so-called vision Zero? And if so, who has a duty to guarantee this right (Nihlén Fahlquist, 2009a; Smids, 2018)? On which philosophical ground is the higher exposure to risk of certain category of road users—vulnerable users but also low-income countries road users—ethically problematic (Nihlén Fahlquist, 2009a; Smids, 2018)? Can there be a right to mobility or accessibility or should accessibility rather be seen as a basic good in Rawlsian terms (Pereira et al., 2017)? What is the best way to incorporate the need of a context-sensitive approach to justice in transportation, Amartya Sen’s capability approach (Pereira et al., 2017) or an approach based on Michael Walzer’s “spheres of justice” (Martens, 2012)? Developing the ethical reflection on these topics is crucial to govern the introduction of new forms of mobility, such as platform-based services like e.g. Uber or vehicle sharing. But it also crucial to rethink the design and regulation of traditional transport systems and to redress the historic inequalities and injustice they embed.

CAVs and road safety: hype and reality

It is somehow interesting that Recommendation 1 of the EC Report on Ethics of CAVs reads “Ensure that CAVs reduce physical harm to persons”. Indeed, the promise of less accidents and more safety is one of the biggest arguments in favour of a strong push to the development of automated driving systems. In public debates over vehicle automation, it is frequent to hear statements such as: The majority of accidents are caused by human errors, computers do not drink and drive, the ratio accident per km of automated vehicles has proven to be way lower than that of human-driven vehicles and so on and so forth. So, to many ears, this recommendation may sound redundant if not misleading. But this is not the case. First of all, one thing is expecting CAVs to reduce accidents, another thing is having solid scientific evidence that this will be the case. Indeed, injury and fatality rates per kilometers are the mainstream metrics of road safety. However, it is at the moment difficult to have reliable statistics about accidents involving CAVs. The accidents are rare also due to the scarcity of CAVs on the road,

not all accidents are registered and even, when they are, details are often missing, including the key information as to whether the automated driving was activated before the crash. It may therefore take decades before we have reliable figures (Kalra & Paddock, 2016). Also, importantly, automated functions and human driving have typically been deployed in different and non-comparable circumstances (Noy et al., 2018), e.g. highways and good weather (CAVs) versus cities and all weathers (human driving); and the role of so-called “human factors”, in particular the potentially unpredictable (new) human-technology interaction is often underestimated in these predictions (Noy et al., 2018). Finally, vehicles equipped with automated features are usually also equipped with the best safety functions, whereas these equipment differ very much across traditional vehicles, so that it may be unclear what the cause of alleged better safety of CAVs is (26).

Another important caveat of this recommendation concerns the distribution of safety across vehicles (different than the distribution across users discussed in recommendation 5 above). Recommendation 6 demands not only to achieve an average decrease in harm across all CAVs or all CAVs of a single developer, but rather a safety improvement for each new model or update of CAV (p. 26). To ensure this, the report recommends that new safety metrics and benchmarks are collaboratively created by policy-makers and researchers, and that manufacturers and deployers are clear about the benchmarks they are using. Also, safety performance should not be assessed once for all, but rather continuously monitored and improved. (p. 26). A related recommendation, though on the data and AI side (and thus included in a different chapter of the report), is that of creating systems for the “auditing” of algorithms (Kroll et al., 2016): “experimental methods for detecting and diagnosing unwanted consequences of algorithmic systems design and operation” (p. 44). Very important is also the creation of a comprehensive, rigorous, and responsible framework for open road testing, possibly at the European level (Recommendation 3, pp. 28–29).

It is in a way telling that the first recommendation of an ethics report concerns an issue, the assessment of future road safety, that many would consider as a merely scientific or technical one. The fact is that in what has been called the age of “post-normal science” (Funtowicz & Ravetz, 1990), the scientific and the political discourse cannot be easily separated. Complex societal challenges like the environment, the digitalization of society, or the response to a pandemic cannot be addressed only by looking at simple facts and predictions. Facts are complex and incomplete, predictions crucially depend on the framework used, policies necessarily require normative and political choices. Future transportation systems will combine two existing complex socio-technical systems—transportation and ICT.

It is therefore very important that “narratives” about possible transportation futures are always critically analysed in relation to different frameworks, and that the scientific and ideological assumptions on which they are grounded are made explicit. A big issue of our time is the extent to which research in digital technologies like AI and its ethics can be expected to be unbiased when performed by scholars or organisations sponsored by private companies who have massively invested in a particular concept of that technology. A similar problem may arise in relation to CAVs and transportation more generally, if the public debates on the future of road safety will be more and more occupied by voices who sell an ideological narrative under the name of “science”. And other scientific, technological and political visions are marginalized or silenced. We might need to turn to a radical interpretation of Responsible Innovation (Blok, 2014), where the dialogue between stakeholders is designed so as to systematically include and empower marginal and radical voices, able to shake deep-rooted convictions of the bearers of dominant narratives.

Safety and freedom

Still under the general topic of road safety, Recommendation 2 proposes to “prevent unsafe use [of CAVs] by inherently safe design”. As obvious as this recommendation may sound to some, this is in fact a very urgent and needed one. Many recent accidents involving automated driving systems, including those which lead to fatal outcomes—such as the Tesla in 2016 and the Uber in 2018—show a recurring pattern. The vehicle encounters a situation that cannot handle, the driver is supposed to intervene, but they fail to do so (in time), leading to a (fatal) crash. Leaving aside the issue of driver’s responsibility, which will be discussed in “Responsibility” section below, it is interesting to note that the riskiness of these handover situations was well-known, and this kind of accidents largely predicted by scientific studies. As reported in the discussion of Recommendation 2, handover scenarios require that the driver is given sufficient time to regain situational awareness (Flemisch et al., 2017), and in one simulator study none of the participants was able to regain control of the car within the 2 seconds they were given to react to a sudden failure in proximity of a curve (Flemisch et al., 2008).

That a feature allowing for- if not inviting dangerous behaviour has been allowed on the road is certainly disturbing—especially considering that automated driving is presented as one possible solution to dangerous driving. However, this is only the last of a long list of failed implementations of technological solutions that could dramatically improve road safety. It is well-known that speeding is one of the main causes of accidents, and still cars are

produced that can go well beyond the speed limits. Intelligent speed adaptation technologies are rarely implemented (Smids, 2018). Drunken driving might be prevented by alcohol interlocks, systems who would require the driver to blow into an in-car breathalyser before starting the vehicle (Grill & Nihlen Fahlquist, 2012; Nihlén Fahlquist, 2009a). As nicely put by the director of the European Traffic Safety Council, we have become so “obsessed with the dream of an autonomous future” that “we are ignoring many of the causes of road collisions that could be avoided today through the use of existing, widely available and affordable technologies” (Avenoso, 2018).

However, this ignoring is far from unintentional and has rather clear ideological roots. The first root is a radical interpretation of the value of individual freedom: “many drivers recoil against anything perceived as a threat to freedom of movement—regardless of the lives that could be saved” (Zipper, 2021). And it matters little that even one of the father of political liberalism, John Stuart Mill, explicitly stated that individual freedom can and should be limited, whenever its exercise directly harms others (Smids, 2018). Another, arguably, more solid theoretical ground for resisting technological limitation to dangerous behaviours on the road is the opposition to “techno-regulation”. This opposition is grounded in the principle of political legitimacy and the rule of law, according to which human behaviour should be influenced by a legitimate political authority via public and predictable norms and not by engineers via changing technological design. It is also grounded in the idea of the intrinsic value of free moral agency, according to which people should be persuaded to comply with moral and legal rules rather than prevented or nudged into compliance. Finally, there is a general concern that delegating the control of human behaviour to scientists and engineers may lead to a dystopian technocratic, brave-new-world society (Smids, 2018). However, while these arguments are certainly relevant in the broader debate on nudging or big data-driven digital technologies (see “[Fairness in data-driven transportation](#)” section below), it is far from obvious that they have the same force in relation to simpler technologies explicitly designed only to prevent one specific illegal and harmful-to-others behaviour, like speeding or drunken driving (Smids, 2018; Yeung, 2011).

Interestingly, in this time of global pandemic many of us have learnt to accept strong limitations to our freedom of movement, in the name of public safety. It has been suggested that there may be lessons to be learnt about road safety in the Covid emergency (Job, 2020). For instance, that protecting the value of human life may require radical political intervention and the rethinking of our lifestyle, or that safety can be achieved only via a combination of technical and political measures. More generally, it is clear that the balancing of public safety and individual freedom in its

various dimensions is yet another topic that a future ethics of transportation will have to put on its agenda.

Meaningful human control

The debate on driver’s freedom and public safety also raises the deeper philosophical question about the meaning of human freedom, and its relation to human control. Consider, again, the recent accidents involving vehicles with automated driving systems discussed above. One way of reading these accidents is indeed that the car manufacturers and regulators decided to protect the driver’s individual freedom by giving them more control of the vehicles, and in this way, they allowed them to endanger the safety of others. However, borrowing the words used by Daniel Dennett in the philosophical debate on free will, the opportunity to interact with a system one doesn’t full understand and cannot easily handle does not seem to be a “variety of free will worth wanting” (Dennett, 1984). In the more recent language of the debate in ethics of technology, it is dubious that these drivers had any “meaningful human control” over their vehicles (Mecacci & Santoni de Sio, 2020; Santoni de Sio & van den Hoven, 2018). The question of control over (automated) vehicles is more complex than it may seems. Being in control does not require just being assigned a task or a role or being in the physical position to perform a certain action. It rather requires having the genuine capacity and opportunity to influence the behaviour of a system, according to one’s own relevant goals, intentions, reasons and values. As it were, sitting in the driver’s seat doesn’t necessarily give you the control of the vehicle. Being in control requires that the technical system responds to one’s relevant reasons, goals or intentions. This is, in a nutshell, the idea of meaningful human control as reason-responsiveness. This idea has important practical implications. First, it allows to see that in cases such as the above-mentioned Tesla and Uber accidents, the driver was not in (meaningful) control of the vehicle, given their arguably limited capacity to interact with the automated system, and therefore to steer it according to their intentions. (Calvert et al., 2020). At the same time, other actors in the chain of design production regulation can be more legitimately deemed in control of the behaviour of the vehicle. The vehicle’s behaviour was indeed more responsive to the higher-level goals, reasons, and capacities of manufacturers and regulators who consciously decide to push automated driving on the road without creating the conditions for a safe interaction with the users (Santoni de Sio & Mecacci, 2021).

Second, the meaningful human control approach clearly shows the need to address the issue of the alignment of the technical system with the capacities of the users, not only via “inherently safe design” (Recommendation 2) but also via

new and better training and educational program for users of new technologies. The latter is explicitly addressed in the EC Report's Recommendation 15 *Promote data, algorithmic, AI literacy and public participation and recommendation* and Recommendation 17 *Promote a culture of responsibility with respect to the obligations associated with CAVs*.

The concept of “meaningful human control” can also help addressing the issue identified by recommendation 4 of the report on traffic rules: *Consider revision of traffic rules to promote safety of CAVs and investigate exceptions to comply with existing traffic rules by CAVs*. Human drivers are sometimes allowed to not follow traffic rules, according to the principle of necessity, when required by exceptional circumstances e.g. when this is the only reasonable way to avoid an accident or a major traffic problem. The report identifies three options for CAVs to be able to not comply with traffic rules when needed: (a) making ex-ante exceptions to some traffic rules for CAVs, (b) having CAVs handing over control to the driver when there may be a need to take a non-compliance decision, or (c) just allowing the CAVs to decide when to not comply, and then assessing retrospectively whether non-compliance was allowed. The report just recommends that policymakers and researchers identify which policy would be more appropriate in which context. In the terms of the theory introduced in this section, this would amount to design the system for meaningful human control. Sometimes we may need to design the system to respond to the practical reasoning of a human person (in the car or in another supervision role), as the person with the best chances of coming to a reasonable judgement about the need for non-compliance. In this case, we need to make sure that they have sufficient time and information to take a responsible decision (p. 30). When this is not possible, the control decision to non-comply may be taken by the vehicle itself. However, this should happen only if this vehicle's decision could demonstrably still be governed by the same principles and reasoning that would govern a responsible human decision. The extent to which machines of the envisageable future will be able to engage in such ethical evaluation is one very hotly debated topic in ethics of AI. Finally, in the light of the difficulty to implement either of these two options, it may be decided to simply change a traffic rule so that CAV can act safely without engaging in non-compliance. In this case it may be said that the control would be shifted even further in the chain, namely to the legislator and, to the extent to which they act in the name of the citizens, to the citizens themselves. In all cases, if designed appropriately, the system would be responding to (different) selected human reasons, and thus being in different ways under the meaningful control of different human agents (Mecacci et al. 2021).

The question of who is ultimately, meaningfully, in control of transportation systems—as opposed to just being in the position to steer vehicles—and how to ensure that

transportation systems are governed by the desired goals and principles, is one who also needs to be addressed by a future ethics (and politics) of transportation. As further explained in the next sections, control is about power and responsibility.

Privacy on the road

Chapter 2 of the EC Report covers a broad set of issues in data and algorithm ethics: privacy, fairness, and explainability. CAVs will probably collect and use “great volumes and varied combinations of static and dynamic data relating to the vehicle, its users, and the surrounding environments” (p. 35). Back in 1995, before the explosion of the data-driven digital technologies and the recent trend of self-driving cars, Jeffrey Reiman published a surprisingly anticipatory article, titled: *Driving to the Panopticon*. Reiman warned that what he then called the “Intelligent Vehicle Highway Systems” may be a massive threat to privacy. Privacy, Reiman argued, results “not only from locked doors and closed curtains, but also from the way our publicly observable activities are dispersed over space and time” (Reiman, 1995, p. 29). In other words, privacy often times depends not as much on information about us not being available, but rather on this information not being available altogether, at the same time, to the same actor. The latter would be what in surveillance studies has become known as the “panopticon effect”.

A related point would be made some year later by the supporters of the idea of “privacy as contextual integrity” (Nissenbaum, 2004). For instance, Michael Zimmer identified what he would call Vehicle Safety Communication Technology as a major domain to which the idea of “privacy in public” should apply. As Zimmer explains, being in a public place does not imply that ‘anything goes’ in terms of collection and storage of personal information. Every arena of life is governed by “norms of information flow”. For instance, just because you are driving on a public road, this doesn't mean that your name or address or profession should be made public to any road users, even though these are all, somehow, “public information”. Even in the public sphere, personal information can be acquired only in certain ways. Similarly, even assuming that someone may have access to your name address etc., i.e. a policewoman who stops you for a routine control, this does not mean that she can share this information with a private company for commercial purposes. Legitimately acquired information is subject also to “norms of distribution”.

Fast forward, beyond 1995 Reiman's “Intelligent Vehicle Highway Systems” and 2005 Zimmer's “Vehicle Safety Communication Technology”, the protection of privacy seems to be an even bigger challenge for CAVs of the 2020s. Data-driven CAV technology can technically be used to

identify passengers through sensors and video monitoring inside the vehicle. It can include personal identification requirements (facial recognition, biometric data, etc.). And this data in conjunction with background information, can be used to develop people's profiles over time (p. 37). "Highly sensitive information about users" can be inferred including their "financial status, ethnicity, political views, personal associations, patterns of habit" (p. 38). This can in turn "have a great impact on the principles of dignity, personal autonomy, and also run against the principles of non-maleficence and justice" (p. 38). One set of recommendations of the report therefore proposes a series of complementary strategies to protect user's privacy. The first and most obvious is the safeguard of informational privacy, in line with some basic principles of GDPR such as data minimization, storage limitation, the strict necessity requirements, as well as the promotion of informed consent practices and user control over data (Recommendation 7).

However, as recognized by the report, user control and consent cannot be the only answer. First, not only users of CAVs will be subject to data collection, but also pedestrians and, potentially, any other road user. These should also be informed and able to protect their privacy as well (Recommendation 10). Relatedly, sensitive data about individuals may be inferred based on (anonymized) data about groups of road users which they happened to be part of (Recommendation 9). Moreover, in order to prevent the above-mentioned panopticon effect, other strategies should be pursued. Competition and consumer law should be leveraged to counteract monopolies of data acquisition on the road and promote user choice and power (Recommendation 8). Finally, people can be able to protect their privacy and resist the power of data-driven systems only if they are equipped with the right technical, educational, motivational, political capacities and opportunities (Recommendation 15).

Once more, ethics of CAVs highlighted an issue, privacy on the road, that deserved to be treated as an important topic of a future ethics of transportation.

Fairness in data-driven transportation

The massive use of data also raises important issues of fairness. The report warns that CAVs market "opens possibilities for differential provision of CAV systems, services and products that pose a risk of perpetuated and increased inequalities between individuals and groups in society" (p. 42). Recommendation 11 warns that the massive introduction of CAVs may lead to persons or groups receiving various forms of "negative special treatments": unequal access to services, discriminatory access, such as deprioritisation in period of high demand, discriminatory differential pricing strategies (p. 44). This can happen either as a "conscious strategy" of

providers and operators or as "unintended consequences of algorithmic bias or biased data in machine-learning models" (44). It can even amount to plain discrimination, if based on features such as "race, gender, social class, income, religion, sexual orientation, place of residence, citizenship, political conviction or religious belief" (44). Solid policies to promote non-discriminatory and more inclusive design are needed.

Also, some data have a clear public value, think of "geographical data, orthographic data, satellite data, weather data, data on crash or near-crash situations including and not including CAVs, and data on mobility, traffic patterns and participants" (46). As key "informational infrastructures and raw material for knowledge and innovation" (47) these data should be kept as much free and open as possible (Recommendation 13), in line with the principles of beneficence and solidarity.

CAVs, in this sense, are not introducing a completely new problem. The introduction of services such as Uber has already clearly showed some risks of data-driven transportation business for transparency and fairness. Presented under the attractive label of "sharing economy", these have often revealed to be aggressive businesses based on a big asymmetry of power between the company running the platform and its users, both drivers and customers (Calo & Rosenblat, 2017). Similar issues may raise with the implementation of mobility as a service. Here is another challenge for a future ethics of transportation: analyzing the extent to which data-driven businesses in transportation can be compatible with the principles of transparency, fairness, beneficence and solidarity.

Responsibility

In 2020, quite at the same time in which the EC Report was published, the US National Transportation Safety Board (NTSB) issued a report on the 2018 UBER accident. Among other things, the report claims that Uber lacked a sufficient "culture of safety" in bringing forward their road tests with self-driving cars and stressed the importance of this culture for the future development of CAVs. The EC Report takes a further step, and in its last chapter 3 "Responsibility", makes a plea for the creation of a "culture of responsibility" around the design and use of CAVs. The chapter argues that safety, as well as justice, freedom, privacy and all the values covered in the report, can only be achieved if different actors involved in the design, development and use of CAVs are aware, able, and motivated to discharge the specific obligations they have in relation to those values.

The EC Report starts from the assumption that responsibility means many different things (Bovens, 1998; Feinberg, 1970; Hart, 1968) and follows (van de Poel & Sand, 2018)

in identifying five different senses of responsibility: two “active” or forward-looking: *obligation* and *virtue*, and three “passive” or backward-looking: *accountability*, *culpability*, and *liability*. The report also focuses on two responsibility issues typical of complex socio-technical systems. The first is the “problem of many hands” (Thompson, 1980; Van de Poel et al., 2015) that is the difficulty of identifying and enforcing individual responsibilities in complex organisations. The second is the “responsibility gap” problem (Matthias, 2004; Santoni de Sio & Mecacci, 2021), that is the further difficulty to identify and enforce individual responsibilities in socio-technical systems involving AI. CAVs systems are arguably potentially affected by both problems. Each sense of responsibility (and its possible gaps) is covered by one specific recommendation.

Recommendation 16 (Responsibility as obligation) starts from the consideration that, as it often happens with new technologies and the social changes they bring, “it can be difficult for manufacturers and deployers, policymakers, users and others to recognise their (new) obligations in relation to the development and use of CAVs” (p. 55). The first task is therefore creating the institutional, social, and educational environment to ensure that all key stakeholders can discuss, identify, decide and accept their respective obligations with respect to CAVs. Manufacturers should become aware and accept their obligation towards safe and inclusive design, policymakers their obligations about safe open road-testing regulations etc. However, as nicely summarized in the opening of Recommendation 17 (Responsibility as Virtue), “knowing your obligations does not amount to being able and motivated to discharge them” (56). If we really want that agents eventually discharge their obligations, we should put them in the conditions to develop the corresponding skills and motivations, that is creating a “culture of responsibility”. This means, among other things, creating spaces to share good practices among manufacturers and policymakers, creating a social and political environment that facilitate and promote the implementations of these good practices, as well as adjusting training and licensing procedures of CAV users.

Moving to backward-looking responsibility, Recommendation 18 (Responsibility as Accountability) focuses on the moral and legal duty to provide an explanation for something that has happened, and one’s role in it. One well-known issue with machine learning system is the “opacity” of their processes, which may make very difficult if not impossible to retrospectively explain the process which lead to a certain outcome. This is known in the literature as the “black box” problem (Castelvecchi, 2016). This problem is usually addressed from a technical point of view, via the development of more “explainable AI”, which is also the focus of Recommendation 14 of the

report. However, in the words of (Noto La Diega, 2018) the difficulty to provide a good explanation for the behaviour of a complex socio-technical systems including AI, may also depend on legal and organizational black boxes, that is legal and organizational frameworks that do not sufficiently allow for- or promote the exchange of meaningful explanations between relevant actors. This means that in addition to have more “explainable” technology, we also need to create the adequate social and legal spaces where questions about the design and use choices about CAVs can be posed and answered; making the relevant people aware, willing and able to provide the required explanations to the relevant audience; and the relevant audiences able and willing to require and understand the explanations. Some of the difficulties in creating the relevant legal and social frameworks to protect the so-called “right to an explanation” for decisions involving automated process are highlighted in the legal and philosophical debate on the GDPR provisions (Edwards & Veale, 2017; Wachter et al., 2017).

Recommendation 19 (Responsibility as Culpability) addresses the often-posed question: Who is to blame (and held legally culpable) for accidents involving CAVs. The report recommends that adequate moral, social, legal, and technical practices are developed to avoid two opposite evils: impunity (or culpability gap), that is no one being legitimately open to moral and/or legal blame for an avoidable accident; and scapegoating, that is some human actors being held “culpable by design”, for instance by being assigned a specific control role or task in the system, without the having a sufficient capacity or opportunity to discharge that role (Elish, 2019). Recommendation 20 invites to also find appropriate legal means to ensure that victims of accidents involving CAVs are compensated.

The different ways in which different transportation systems may create or worsen different responsibility gaps is yet another topic to be included in an ethics of transportation beyond CAVs. Responsibility gaps, both in the more traditional version of the problem of many hands, and in their more recent versions created by AI, are very likely of happen in complex socio-technical systems powered by AI like many future transportation systems will be. Think of the collapse of the Morandi bridge in Italy or the accidents involving the Boeing 737 Max as just two recent tragic examples. As also suggested by the report in relation to CAVs, one model to looked at is indeed that of civil aviation, where high level of complexity and automation have been coupled with equally high levels of safety, control, and responsibility. Thanks to a comprehensive design approach, combining technical, organizational, socio-psychological and legal elements.

Conclusion and future perspectives

This paper has a big ambition and a limited scope. The big ambition is to contribute to the establishment of ethics of transportation as an independent domain of applied ethics. Ethics of transportation has been defined as the discipline engaged in a systematic reflection on the values, principles, norms, and concepts that should guide the design, production, regulation, and use of different *transportation systems and infrastructures* in different contexts. It is narrower than ethics of mobility insofar as it does not covers issues such as the right to freely move across countries and ethics of tourism. But it is broader than a professional ethics for transportation engineers, insofar as it also concerns the role of policy-makers, users, and persons more generally in the design, regulation and use of transportation systems. It is also broader than ethics of CAVs or automated transportation. Whereas there do exist some important contributions to ethics of transportation, especially in relation to issue of distributive justice, a full-fledged discipline systematically covering a broader range of topics such as that presented in this paper is still missing. The paper has presented the main results of the report *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility* written by the Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659) (Bonnefon et al., 2020). In the original report, the twenty recommendations are presented in three chapters—road safety, data and algorithm ethics, and responsibility. This paper has discussed them according to eight philosophically topics: Responsible Innovation, road justice, road safety, freedom, meaningful human control, privacy, data fairness, responsibility. These topics are also proposed as a starting point for a future ethics of transportation.

However, the scope of this paper has been limited. First, it only covers some but not all possible topics in ethics of road transportation. Further research will have to identify and discuss other topics. Environmental sustainability and the future of work are just two examples of important topics not covered in this paper. Second, the paper does not cover ethical issues related to domains of transportation other than road: railway, aviation, sea, space. These will also have to be included in a future ethics of transportation. Finally, this paper has taken a theoretical, philosophical perspective on ethics of transportation. However, all concepts and ideas in ethics of transportation will have to be discussed and adjusted also according to an interdisciplinary and applied approach. Experts from different disciplines will have to clarify and operationalized them in relation to specific domains of applications and through a

close interaction between technical, philosophical, social, political, legal perspective. There is a long road ahead of ethics of transportation.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Avenoso, A. (2018). The threat to progress on preventing road deaths. *Financial Times*.
- Bellon, T. (2018). *Fatal U.S. self-driving auto accident raises novel legal questions*. Reuters. <https://www.reuters.com/article/us-autos-selfdriving-uber-liability-anal/fatal-u-s-self-driving-auto-accident-raises-novel-legal-questions-idUSKBN1GW2SP>
- Blok, V. (2014). Look who's talking: Responsible innovation, the paradox of dialogue and the voice of the other in communication and negotiation processes. *Journal of Responsible Innovation*, 1(2), 171–190. <https://doi.org/10.1080/23299460.2014.924239>
- Bonnefon, J.-F., Černý, D., Danaher, J., Devillier, N., Johansson, V., Kovacikova, T., Martens, M., Mladenović, M., Palade, P., Reed, N., Santoni de Sio, F., Tsinorema, S., Wachter, S., & Zawieska, K. (2020). *Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: Recommendations on road safety, privacy, fairness, explainability and responsibility*. Publication Office of the European Union: Luxembourg.
- Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576. <https://doi.org/10.1126/science.aaf2654>
- Bovens, M. (1998). *The quest for responsibility: Accountability and citizenship in complex organisations*. Cambridge University Press.
- Calo, R., & Rosenblat, A. (2017). The taking economy: Uber, information, and power. *Columbia Law Review*, 117(6), 1623–1690. <https://doi.org/10.2139/ssrn.2929643>
- Calvert, S. C., Mecacci, G., van Arem, B., Santoni de Sio, F., Heikoop, D. D., & Hagenzieker, M. (2020). Gaps in the control of automated vehicles on roads. *IEEE Intelligent Transportation Systems Magazine*. <https://doi.org/10.1109/ITS.2019.2926278>
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, 538(7623), 20–23. <https://doi.org/10.1038/538020a>
- Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. MIT Press.
- Di Fabio, U., Broy, M., & Brügger, R. J. (2018). *Task force on ethical aspects of connected and automated driving*. Federal Ministry of Transport and Digital Infrastructure of the Federal Republic of Germany.
- Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a right to explanation is probably not the remedy you are looking for. *Duke Law and Technology Review*, 16(1), 1–65. <https://doi.org/10.2139/ssrn.2972855>

- Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human–robot interaction. *Engaging Science, Technology, and Society*, 5, 40. <https://doi.org/10.17351/ests2019.260>
- European Commission. (2019). *High-level expert group on artificial intelligence Ethics Guidelines For Trustworthy AI*.
- European Group on Ethics in Science and New Technologies. (2018). *Artificial intelligence, robotics and 'Autonomous' systems*.
- Feinberg, J. (1970). *Doing & deserving: essays in the theory of responsibility*. Princeton University Press.
- Flemisch, F., Altendorf, E., Canpolat, Y., Weßel, G., Baltzer, M., Lopez, D., Herzberger, N. D., Voß, G. M. I., Schwalm, M., & Schutte, P. (2017). Uncanny and unsafe valley of assistance and automation: First sketch and application to vehicle automation. In C. M. Schlick, S. Duckwitz, F. Flemisch, M. Frenz, S. Kuz, A. Mertens, & S. Mütze-Niewöhner (Eds.), *Advances in ergonomic design of systems, products and processes* (pp. 319–334). Springer. https://doi.org/10.1007/978-3-662-53305-5_23
- Flemisch, F., Kelsch, J., Löper, C., Schieben, A., Schindler, J., & Heesen, M. (2008). Cooperative control and active interfaces for vehicle assistance and automation. In *Cooperative Control and Active Interfaces for Vehicle Assistance and Automation. FISITA World Automotive Congress, Munich*.
- Funtowicz, S. O., & Ravetz, J. R. (1990). Post-normal science: A new science for new times. *Scientific European*, 266, 20–22.
- Goodall, N. J. (2016). Away from trolley problems and toward risk management. *Applied Artificial Intelligence*, 30(8), 810–821. <https://doi.org/10.1080/08839514.2016.1229922>
- Grill, K., & NihlenFahlquist, J. (2012). Responsibility, paternalism and alcohol interlocks. *Public Health Ethics*, 5(2), 116–127. <https://doi.org/10.1093/phe/phis015>
- Hansson, S. O. (2014). Making road traffic safer: Reply to ori. *Philosophical Papers*, 43(3), 365–375. <https://doi.org/10.1080/05568641.2014.976439>
- Hart, H. L. A. (1968). *Punishment and responsibility*. Oxford University Press.
- Hevelke, A., & Nida-Rümelin, J. (2014). Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-014-9565-5>
- Himmelreich, J. (2019). Ethics of technology needs more political philosophy. *Communications of the ACM*, 63(1), 33–35. <https://doi.org/10.1145/3339905>
- Hosmer, L. T. (1996). The call for transportation ethics. *Transportation Quarterly*, 50(1), 22–33.
- Husak, D. (2004). Vehicles and crashes: Why is this moral issue overlooked? *Social Theory and Practice*, 30(3), 351–370.
- JafariNaimi, N. (2018). Our bodies in the trolley's path, or why self-driving cars must *not* be programmed to kill. *Science, Technology, & Human Values*, 43(2), 302–323. <https://doi.org/10.1177/0162243917718942>
- Jeekel, H. (2018). *Inclusive transport: Fighting involuntary transport disadvantages*. Elsevier.
- Job, S. (2020). Can COVID-19 teach us something for the road safety epidemic? *Transport for Development*. <https://blogs.worldbank.org/transport/can-covid-19-teach-us-something-road-safety-epidemic>
- Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, 94, 182–193. <https://doi.org/10.1016/j.tra.2016.09.010>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. *University of Pennsylvania Law Review*, 165, 49.
- Liability for Artificial Intelligence and other emerging digital technologies*. (2019).
- Lin, P. (2014). The robot car of tomorrow may just be programmed to hit you. *Wired*.
- Lin, P. (2015). Why Ethics Matters for Autonomous Cars. In *Autonomes Fahren* (pp. 69–85). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45854-9_4
- Liu, H.-Y. (2017). Irresponsibilities, inequalities and injustice for autonomous vehicles. *Ethics and Information Technology*, 19(3), 193–207. <https://doi.org/10.1007/s10676-017-9436-2>
- Manders-Huits, N. (2011). What values in design? The Challenge of incorporating moral values into design. *Science and Engineering Ethics*, 17(2), 271–287. <https://doi.org/10.1007/s11948-010-9198-2>
- Martens, K. (2012). Justice in transport as justice in accessibility: Applying Walzer's 'Spheres of Justice' to the transport sector. *Transportation*, 39(6), 1035–1053. <https://doi.org/10.1007/s11116-012-9388-7>
- Martens, K. (2017). *Transport justice*. Routledge.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- Mecacci, G., & Santoni de Sio, F. (2020). Meaningful human control as reason-responsiveness: The case of dual-mode vehicles. *Ethics and Information Technology*, 22(2), 103–115. <https://doi.org/10.1007/s10676-019-09519-w>
- Mecacci G., Calvert S., & Santoni de Sio F. (forthcoming 2021). *Human-machine coordination in mixed traffic as a problem of Meaningful Human Control*. AI and Society.
- Millar, J. (2014a). An ethical dilemma: When robot cars must kill, who should pick the victim? *Robohub*.
- Millar, J. (2014b). Proxy prudence: Rethinking models of responsibility for semi-autonomous robots. *We Robot Conference*. <https://doi.org/10.2139/ssrn.2442273>
- Mindell, D. A. (2015). *Our robots, ourselves: Robotics and the myths of autonomy*. Penguin.
- Mladenović, M. N. (2017). Transport justice: Designing fair transportation systems. *Transport Reviews*, 37(2), 245–246. <https://doi.org/10.1080/01441647.2016.1258599>
- NihlenFahlquist, J. (2009a). Saving lives in road traffic—ethical aspects. *Zeitschrift Fur Gesundheitswissenschaften = Journal of Public Health*, 17(6), 385–394. <https://doi.org/10.1007/s10389-009-0264-7>
- NihlenFahlquist, J. (2009b). Saving lives in road traffic—Ethical aspects. *Zeitschrift Fur Gesundheitswissenschaften*, 17(6), 385. <https://doi.org/10.1007/s10389-009-0264-7>
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119–157.
- Norton, P. D. (2008). *Fighting traffic: The dawn of the motor age in the American City*. MIT Press.
- Noto La Diega, G. (2018). Against the dehumanisation of decision-making. Algorithmic decisions at the crossroads of intellectual property, data protection, and freedom of information. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*. <https://doi.org/10.31228/osf.io/s2jnk>
- Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, 102, 68–78. <https://doi.org/10.1016/j.ssci.2017.07.018>
- Nyholm, S. (2018a). The ethics of crashes with self-driving cars: A roadmap. I. *Philosophy Compass*, 13(7), e12507. <https://doi.org/10.1111/phc3.12507>
- Nyholm, S. (2018b). The ethics of crashes with self-driving cars: A roadmap. II. *Philosophy Compass*, 13(7), e12506. <https://doi.org/10.1111/phc3.12506>
- Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical Theory and Moral Practice*, 19(5), 1275–1289. <https://doi.org/10.1007/s10677-016-9745-2>

- On the road to automated mobility: An EU strategy for mobility of the future.* (2018).
- Ori, M. (2014). The morality of motorcycling. *Philosophical Papers*, 43(3), 345–363. <https://doi.org/10.1080/05568641.2014.976438>
- Ori, M. (2020). Why not road ethics? *Theoria*, 86(3), 389–412. <https://doi.org/10.1111/theo.12248>
- Pereira, R. H. M., Schwanen, T., & Banister, D. (2017). Distributive justice and equity in transportation. *Transport Reviews*, 37(2), 170–191. <https://doi.org/10.1080/01441647.2016.1257660>
- Rajan, S. C. (2007). Automobility, liberalism, and the ethics of driving. *Environmental Ethics*, 29(1), 77–90. <https://doi.org/10.5840/enviroethics200729130>
- Reiman, J. H. (1995). Driving to the panopticon: A philosophical exploration of the risks to privacy posed by the highway technology of the future. *Computer & High Technology Law Journal*, 11(1), 27–44.
- Richardson, B. C. (1995). Transportation ethics. *Transportation Quarterly*, 49(2), 117.
- Sager, T., & Bergmann S. (Eds.) (2016). *The ethics of mobilities*. Routledge. <https://doi.org/10.4324/9781315616186>.
- Santoni de Sio, F. (2016). *Ethics and self-driving cars: A white paper on responsible innovation in automated driving systems*. Dutch Ministry of Infrastructure and Environment Rijkswaterstaat.
- Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy and Technology*. <https://doi.org/10.1007/s13347-021-00450-x>
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*. <https://doi.org/10.3389/frobt.2018.00015>
- Smids, J. (2018). The moral case for intelligent speed adaptation. *Journal of Applied Philosophy*, 35(2), 205–221. <https://doi.org/10.1111/japp.12168>
- Stilgoe, J. (2020). Who killed Elaine Herzberg? In J. Stilgoe (Ed.), *Who's Driving Innovation?* (pp. 1–6). Springer International Publishing. https://doi.org/10.1007/978-3-030-32320-2_1
- Stilgoe, J. (2017). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science*. <https://doi.org/10.1177/0306312717741687>
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- Task force on ethical aspects of connected and automated driving. (2018). *Report*.
- Thompson, D. F. (1980). Moral responsibility of public officials: The problem of many hands. *The American Political Science Review*, 74(4), 905–916.
- Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 94(6), 1395–1415. <https://doi.org/10.2307/796133>
- Van de Poel, I., Royakkers, L. M. M., & Zwart, S. D. (2015). *Moral responsibility and the problem of many hands*. Routledge. <https://doi.org/10.4324/9781315734217>
- Van de Poel, I., & Sand, M. (2018). Varieties of responsibility: Two problems of responsible innovation. *Synthese*. <https://doi.org/10.1007/s11229-018-01951-7>.
- Van den Hoven, J., Lokhorst, G. J., & Van de Poel, I. (2012). Engineering and the problem of moral overload. *Science and Engineering Ethics*, 18(1), 143–155. <https://doi.org/10.1007/s11948-011-9277-z>
- Van Wee, B. (2011). *Transport and ethics*. Elgar Publishing.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.
- Yadron, D., & Tynan, D. (2016). *Tesla driver dies in first fatal crash while using autopilot mode* [Technology] *The Guardian*. <https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk>
- Yeung, K. (2011). Can we employ design-based regulation while avoiding brave new world? *Law, Innovation and Technology*, 3(1), 1–29. <https://doi.org/10.5235/175799611796399812>
- Zipper, D. (2021). *The life-saving car technology no one wants*. Bloomberg City Lab.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.