AI & SOCIETY (2024) 39:669–673
https://doi.org/10.1007/s00146-022-01495-y

OPEN FORUM

A new standard for accident simulations for self-driving vehicles: Can we use Waymo's results from accident simulations?

Björn Lundgren[1,2,3]

Received: 17 February 2022 / Accepted: 28 March 2022 / Published online: 21 June 2022

Abstract
Recent simulations by Scanlon et al. showed seemingly spectacular results for the Waymo self-driving vehicle in simulations of real accident situations. In this paper, it is argued that the selection criteria for accident situations must be modified in accordance with the relevant policy alternatives. While Scanlon et al. compare Waymo with *old* human-driven vehicles, it is argued here that the relevant policy question is whether we ought to use self-driven vehicles or human-driven vehicles *in the future*, which means that we need to consider whether other technological solutions, which are available but not broadly used in human-driven vehicles, could result in human-driven vehicles managing to avoid the same accidents. In this article, a proposal for a new standard of selection criteria is made.

Keywords Selection criteria · Accident simulations · Self-driving vehicles · Autonomous vehicles · Human-driven vehicles · Policy considerations

1 Introduction

How should we evaluate the safety of autonomous vehicles? Various methods have been suggested, such as observational data, simulations, mathematical proofs, and tests similar to human tests for driving licenses (see, e.g., Hicks 2018; Lundgren 2021; McBride 2016). Yet, none of these options are without problems (see Hicks 2018; Lundgren 2021 for critical discussions). Recently, Scanlon et al. (2021) reported on simulation results for Waymo's self-driving vehicle, using a new simulation method with seemingly spectacular results. The simulation method tested how the Waymo vehicles would behave in real accident situations (based on examples from accidents involving human-driven vehicles). When the Waymo driver replaced the driver that caused the crash, they avoided all accidents. In the remaining cases (i.e., when the Waymo driver replaced the driver that was hit), they avoided 82% of the accidents and deemed that they mitigated the crash severity in 10% of the cases.

The fact that Waymo, in simulation, managed to avert most human-caused accidents might seem very impressive; indeed, it is a positive result. However, setting aside the methodological limitation of simulation results from self-driving vehicles (cf. Hicks 2018; Lundgren 2021), I will suggest that we need a new standard of criteria for selection data for accident simulations for self-driving vehicles—according to which the results from Scanlon et al. would not be valid.[1] In the remainder of this article, I will make two suggestions for improvement. First, I will suggest a complementary simulation that takes into consideration that the type of errors that a machine makes may differ from the type of errors that humans make (Lundgren 2021). Second, I will discuss the selection criteria used by Scanlon et al., to argue that we should not consider accidents that can be avoided or mitigated by other technologies (such as alcohol locks and speeding controls). This is motivated by policy considerations according to which the permissibility of wide-scale implementation of self-driving vehicles *prima facie* depends on them being at least as safe as human-driven vehicles (Hicks 2018; Lundgren 2021). Moreover, if

✉ Björn Lundgren
b.a.lundgren@uu.nl; bjorn.lundgren@iffs.se; bjorn.lundgren@philosophy.su.se

[1] Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands

[2] Institute for Futures Studies, Stockholm, Sweden

[3] Department of Philosophy, Stockholm University, Stockholm, Sweden

[1] Scanlon et al. (2021, p. 2) recognizes that safety must be evaluated by a multitude of methodologies referring to previous work in Webb et al. (2020).

the safety of human-driven vehicles can be technologically enhanced, then self-driving vehicles should be compared with the improved vehicles.[2] Based on the aforementioned principle—which we can call *the safety criterion*—I will argue that we need to alter the selection criteria for the type of accident simulations that Waymo has performed.[3]

Lastly, while there is a large set of policy questions that must be answered in order to determine whether self-driving vehicles ought to be implemented, most of those questions are beyond the scope of this paper. What I focus on here is the policy question directly related to the work of Scanlon et al. (i.e., questions about safety). However, I will return to briefly comment on some of the broader policy concerns in the concluding section (see Sect. 3). For some of the relevant policy questions see the ethical overview by Hansson et al. (2021).

2 The two problems

In this section, I will describe the two problems of the methods used by Scanlon et al. First, a fundamental problem for the method of accident evaluations of self-driving vehicles that Scanlon et al. applies is that although such vehicles may avoid human-made accidents, this does not necessarily say anything about the vehicles' actual success-rate. Simply put, machines may make mistakes that humans do not make and therefore an appropriate comparison must include machines possibilities for errors rather than merely looking at situations in which humans fail. More specifically, machines' traffic behavior might differ from humans' traffic behavior, as is also recognized by Scanlon et al. (2021, p. 13; cf. Lundgren 2021). This is illustrated by the infamous Tesla death accident in which the autopilot tried to drive under a trailer "with the bottom of the trailer impacting the windshield of the Model S",[4] a choice no rational human would make (plausibly, the human driver was not actively controlling the vehicle). Hence, even if machines would avoid all human-type accidents, that does not mean that they would not cause other types of accidents.

One way to deal with the above-mentioned problem would be to simulate how human drivers would behave in cases when self-driving vehicles fail. If the Tesla example is illustrative of the type of errors that self-driving vehicles make, it might turn out that human drivers are equally successful in avoiding machine-cause accidents as Waymo was in avoiding human-caused accidents.

Second, recognizing that some human pre-crash behavior would differ from how the Waymo Driver would perform in a similar situation, Scanlon et al. opted to align the process. As they noted, the human driver might have been speeding and Waymo is designed not to exceed the speed limit. So, they aligned

> the Waymo Driver trajectory with the human-driven trajectory so that, absent any collision avoidance behavior, a similar collision scenario as was experienced by the original human driver would occur. (Scanlon et al. 2021, p. 6)

However, as I will argue, this alignment process is not the best approach. Consider the fact that human errors are the major cause of accidents (in the USA approximately 93% of accidents are human caused). Of the human-caused accidents, a third are due to intoxication, 30% to speeding, and 20% to distracted drivers (Fagnant and Kockelman 2015).

Unsafe human behavior might seem like a good reason to think that self-driving vehicles are safer than human-driven vehicles. Simply put, self-driving vehicles do not drink, do not speed, and do not get distracted in the same way as humans do (Lundgren 2021, p. 412). However, it is also possible to use technology to minimize or completely remove all of these main causes: alcohol interlocks, smart speed controls, and most recently AI-technology that evaluates human focus. (ibid; Nyholm and Smids 2020, p. 339).

Thus, contrary to Scanlon et al.'s alignment process, the lesson we should draw from Lundgren's argument is *not* that accident scenarios should be aligned, but rather that many of the accidents that were aligned should have been *excluded* from a reasonable comparison. The reason for that is simple: technologies allow us to develop better human-driven (technologically assisted) vehicles and we should compare self-driving vehicles against "accident rates for *future* human-driven vehicles, not accidents rates using old technology" (Lundgren 2021, p. 412). Why is that? Simply put, the policy question we are facing when we ask whether self-driving vehicles should be allowed is not simply whether they are better than *old* human-driven vehicles but whether the vehicles replacing the old human-driven vehicles should be self-driving or technically enhanced human-driven vehicles

[2] It is worthwhile pointing out that the distinction between human-driven vehicles and self-driving vehicles is not binary. Indeed, the different levels of autonomy (well-known from, e.g., SAE 2018) illustrate this non-binary nature. The improved human-driven vehicles that I will use as an example may be considered as level 1 (driver assisted), while Scanlon et al. talk of level 4. Only level 5 is fully automatic in all contexts.

[3] Hicks (2018) talks of the *safety argument* as an argument that promotes self-driving vehicles because they will be safer than human-driven vehicles. I talk of the *safety criterion* as a *prima facie* principle because it may—in the face of new evidence for policy-considerations—be overridden. For example, if there are other benefits of self-driving vehicles, then that may override a requirement that self-driving vehicles should be as safe as the best kind of human-driven vehicle. This is especially important if these are lifesaving and thus can reduce the overall effect of harm from the traffic system (see Sect. 3 for a slightly more detailed discussion).

[4] See: https://www.tesla.com/blog/tragic-loss.

(e.g., including alcohol locks, speed controls, and technology that ensures that the driver is focused). That is, the question we are addressing is a policy question regarding what kind of vehicles should be implemented into the transport system as the current fleet of vehicles are being replaced. If human-driven vehicles (with appropriate technical updates) are safer than self-driving vehicles, then, *ceteris paribus*, the current fleet of vehicles should be replaced by technically enhanced human-driven vehicles (including alcohol locks, speed controls, and systems to ensure that the driver is focused), not self-driving vehicles. Hence, we need to evaluate self-driving vehicles in light of other alternatives, not in light of the current fleet of vehicles. (*N.B.*, although I talk of future human-driven vehicles, the example I give is based on already available technology.)[5]

Based on this I suggest that accident evaluation for self-driving vehicles of the kind used by Waymo should be jettisoned and replaced by—or, at the very least complemented by—accident evaluations that *only* include accidents that cannot be avoided or mitigated with alcohol locks, speeding controls, and smart devices to ensure driver attention.[6] This would allow for a better comparison between the future offered by self-driving vehicles and the future offered by human-driven vehicles. The aim should be to evaluate self-driving vehicles' success in situations when human drivers fail due to the *essential* limitations of human driving, not due to the limitation of improper human driving behavior when such behavior can be avoided, or severally mitigated, using technological safeguards. Moreover, in cases where self-driving vehicles cause accidents, we should check whether humans' driving behavior would have avoided the accident. Combining these two test ideas would yield more fruitful and policy-relevant data about the safety of self-driving vehicles as compared with the best human-driven vehicles.

Lastly, what I have argued here may be generalized to other alternatives. That is, there may be more possible alternatives for which types of vehicles should replace the current fleet of vehicles. For example, Müller and Gogoll (2020) discuss the option of AI-assisted human-driven vehicles. Which alternatives need to be considered is a complicated question. However, that does not affect the argument I am making here: we need to compare the success of accident avoidance of self-driving vehicles against other alternatives for replacing the current fleet of vehicles, not against the current fleet of vehicles.

3 Concluding discussion

I have argued that the evaluation method promoted by Scanlon et al. must be modified in order to respond to the policy question posed by the safety criterion (i.e., whether self-driving vehicles are safe enough to be permissible), because when we evaluate traffic safety in general, or accident avoidance in particular, we need to keep in mind what policy problem such scientific questions should be responding to. In this case it is clear that we are talking about the future of transportation. Hence, we cannot compare the success of self-driving vehicles against old technology. Rather we ought to compare it to the alternative, which minimally should include human-driven vehicles enhanced with accident-avoidance technology (such as speed controls, alcohol locks, and smart technology that ensures that the driver is focused).

However, in response to these arguments, one may worry about the extent to which we can have knowledge about other future alternatives. As I argued earlier, we need not worry about this, because I have considered options that are technically available but not broadly used. So the situation is quite similar to the self-driving vehicles we want to compare with (see fn. 5).

Moreover, the discussion here does not correspond to the complexity of all policy considerations that concern choices about future traffic systems. If we recognize the limits of Scanlon et al. in relation to *some* policy considerations, then should we not address *all* plausible policy considerations? As noted earlier, the safety criterion can be overridden and there may be other considerations that turn out to be more important in the policy choice of which means of transportation that ought to permissible, impermissible, benefitted, or enforced in the future. However, if we broaden the analysis then we must also recognize that there are many questions that remain unresolved (see, e.g., Hansson et al. 2021 for an overview). Nevertheless, if we are addressing safety concerns, then it seems reasonable to limit the relevant policy questions specifically to those that concern safety.

Yet, even if we limit the policy debate to the safety criterion, there is still a broader set of issues that could be considered. Take, for example, the vision zero traffic policy (see, e.g., Belin et al. 2012). According to this policy the aim is to ensure that the traffic system yields no fatalities or

[5] I make this note in response to the worry that information about *future* human-driven vehicles may be epistemically inaccessible. But the point of comparison that I consider in this article is adding technology that is available but not broadly used. Although we might need to calculate a failure rate for alcohol locks, speed controls, or focus-controls, based on available data, that is not different from the situation of self-driving vehicles.

[6] Although my point here is that the results from Scanlon et al. are not useful for the relevant policy questions, I do not want to say that their method is useless. It still provides some information and the method can provide an interesting point of comparison between different self-driving vehicles (even if the point of comparison should not, as I argue, be old human-driven vehicles). My point is merely that in order to satisfy the most relevant policy question (i.e., what type of vehicles should replace the current fleet of vehicles in the future), we need to use a different set of data.

serious injuries. However, traffic accidents are not the only cause of fatalities or serious injuries from the traffic system. They are also caused by climate, environmental, and health effects, due to the combustion engine's exhaust, for example. Could self-driving vehicles perform better than the alternatives relative to the metrics for some of these issues? Most studies show that self-driving vehicles will *increase* traffic (see, e.g., Pernestål and Kristoffersson 2019; Soteropoulos et al. 2018), which might favor human-driven vehicles unless the self-driving vehicles are complemented by regulations that increase efficiency.[7] Moreover, given the earlier constraint that we ought to compare future options against other future options, we cannot suppose any benefits from the fact that self-driving vehicles are normally electric (for the other future alternatives should be human-driven electric vehicles). Furthermore, if we were to broaden this analysis, then it would be preferable if traffic from driven vehicles (self-driven or human-driven) were replaced by traffic from man-powered means of transportation (such as bicycles). The reasons for that are simple. First, as already implied, it would save lives through reducing climate, environmental, and negative health impacts due to pollution and production. Second, we can save many lives through the benefits of exercise (see, e.g., Sommar et al. 2021 for a case-study).

However, although these factors are important from the perspective of a broader policy debate, they do not seem to affect the point I am making in this article. That is, there is no reason to think that the methodological constraints on simulations that follow from adhering to the safety criterion limit scientific evaluations in a way that is contrary to the appropriate policy considerations. In fact, the suggestions I make seem to be in line with an overall argument that we need to evaluate future technological option in relation to all relevant alternatives.

Acknowledgements I am grateful for the review comments I received. I also want to thank Sven Nyholm and Karoliina Pulkkinen for comments I received prior to submission.

Funding This work is part of the research program Ethics of Socially Disruptive Technologies, which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

Data availability The manuscript has no associated data.

[7] For example, co-travelling and policies that promote taxi-services rather than private ownership (i.e., so that the total vehicle fleet can be reduced).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Belin MÅ, Tillgren P, Vedung E (2012) Vision zero—a road safety policy innovation. Int J Inj Contr Saf Promot 19(2):171–179. https://doi.org/10.1080/17457300.2011.635213

Fagnant DJ, Kockelman K (2015) Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transport Res Part A 77:167–181. https://doi.org/10.1016/j.tra.2015.04.003

Hansson SO, Belin MÅ, Lundgren B (2021) Self-driving vehicles—an ethical overview. Philos Technol 34:1383–1408. https://doi.org/10.1007/s13347-021-00464-5

Hicks DJ (2018) The safety of autonomous vehicles: lessons from philosophy of science. IEEE Technol Soc Mag 37(1):62–69. https://doi.org/10.1109/MTS.2018.2795123

Lundgren B (2021) Safety requirements vs. crashing ethically: what matters most for policies on autonomous vehicles. AI Soc 36:405–415. https://doi.org/10.1007/s00146-020-00964-6

McBride N (2016) The ethics of driverless cars. ACM SIGCAS Comput Soc 45(3):179–184. https://doi.org/10.1145/2874239.2874265

Müller JF, Gogoll J (2020) Should manual driving be (eventually) outlawed? Sci Eng Ethics 26:1549–1567. https://doi.org/10.1007/s11948-020-00190-9

Nyholm S, Smids J (2020) Automated cars meet human drivers: responsible human-robot coordination and the ethics of mixed traffic. Ethics Inf Technol 22:335–344. https://doi.org/10.1007/s10676-018-9445-9

Pernestål A, Kristoffersson I (2019) Effects of driverless vehicles–comparing simulations to get a broader picture. Eur J Transport Infrastruct Res 19(1):1–23. https://doi.org/10.18757/ejtir.2019.19.1.4079

SAE (2018) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Ground vehicle standard J3016_201806. SAE Int. https://doi.org/10.4271/J3016_201806

Scanlon JM, Kusano KD, Daniel T, Alderson C, Ogle A, Victor T (2021) Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. Accid Anal Prev 163:106454. https://doi.org/10.1016/j.aap.2021.106454

Sommar JN, Johansson C, Lövenheim B, Schantz P, Markstedt A, Strömgren M, Stigson H, Forsberg B (2021) Overall health impacts of a potential increase in cycle commuting in Stockholm, Sweden. Scandinavian J Public Health. https://doi.org/10.1177/14034948211010024

Soteropoulos A, Berger M, Ciari F (2018) Impacts of automated vehicles on travel behaviour and land use: an international review of modelling studies. Transp Rev 39(1):29–49. https://doi.org/10.1080/01441647.2018.1523253

Webb N, Smith D, Ludwick C, Victor T, Hommes Q, Favaro F et al (2020) Waymo's safety methodologies and safety readiness determinations. arXiv preprint. https://arxiv.org/abs/2011.00054.