ORIGINAL PAPER

Automated vehicles and the morality of post-collision behavior

Sebastian Krügel[1,2] · Matthias Uhl[1,2] · Bryn Balcombe[3]

Abstract

We address the considerations of the European Commission Expert Group on the ethics of connected and automated vehicles regarding data provision in the event of collisions. While human drivers' appropriate post-collision behavior is clearly defined, regulations for automated driving do not provide for collision detection. We agree it is important to systematically incorporate citizens' intuitions into the discourse on the ethics of automated vehicles. Therefore, we investigate whether people expect automated vehicles to behave like humans after an accident, even if this behavior does not directly affect the consequences of the accident. We find that appropriate post-collision behavior substantially influences people's evaluation of the underlying crash scenario. Moreover, people clearly think that automated vehicles can and should record the accident, stop at the site, and call the police. They are even willing to pay for technological features that enable post-collision behavior. Our study might begin a research program on post-collision behavior, enriching the empirically informed study of automated driving ethics that so far exclusively focuses on pre-collision behavior.

Keywords Post-collision behavior · Detection of accidents · Accident data · Trolley · Experiment

Introduction

Ethical considerations regarding automated driving struggle to keep pace with technological developments. Thus, the European Commission Expert Group's (Horizon, 2020) recommendations on the ethics of connected and automated vehicles (AVs) prove valuable for this crucial endeavor. The panel's report concerns many ethically relevant issues, such as traffic safety, risk distribution, data processing, privacy, and questions of responsibility for adverse events. The report also recommends that authorities and researchers provide collision and near-collision data for accident investigations, although the recommendation's scope does not allow for details. However, the report does call for a dialogue among policymakers, manufacturers, deployers of AVs, and other stakeholders to specify what data are particularly valuable in the context of AV collision or near-collision events and how these data should be made available.

In the present study, we address the report's considerations for data provision in the event of collisions and further include appropriate post-collision behaviors of AVs in general. Although post-collision behavior (i.e., what a human driver should do in case of an accident) is an essential aspect of road traffic, this issue is almost never contemplated regarding AVs, possibly because people expect (and demand) AVs to significantly increase road safety. However, no one expects zero accidents once AVs are allowed on the road, and certainly not while they operate in mixed traffic together with human-controlled cars (e.g., Goodall, 2014, 2016a; Nyholm & Smids, 2016). It is therefore important to define how AVs should behave when accidents inevitably occur.

Human drivers' appropriate post-collision behavior is clearly defined. Article 31 of the 1968 Convention on Road Traffic (UNECE, 2021) specifies that, in case of an accident, the (human) driver must stop at the accident site, alert the hazard to other road users, contact emergency services, and help determine responsibilities, possibly simply by preserving traces of the accident. In many countries, a driver's failure to execute these basic post-collision behaviors is a criminal offence. Nonetheless, over 700,000 hit-and-run

✉ Matthias Uhl
matthias.uhl@thi.de

1 Faculty of Computer Science, Technische Hochschule Ingolstadt, Ingolstadt, Germany

2 School of Governance, Technical University of Munich, Munich, Germany

3 Autonomous Drivers Alliance (ADA), London, UK

accidents are estimated to have occurred in 2015 in the United States alone (AAA Foundation, 2018). The question is whether AVs will significantly reduce the high proportion of hit-and-run accidents. So far, eCall and Event Data Recorders exist for occupant protection and will not trigger for collisions with vulnerable road users. Moreover, recent UNECE regulations for AVs' "black-box recorders" include no provision for collision detection. Last, but not least, a 2020 amendment proposal from the Convention on Road Traffic suggests exempting automated driving systems from 32 articles and 132 clauses related to driver behavioral requirements, including Article 31 (Economic Commission for Europe, 2020).

It is unlikely that the automotive industry will proactively close the regulatory gap caused by the amendments to the Convention on Road Traffic because the topic of AVs' collision detection and appropriate post-collision behavior is so far largely ignored by scientists, the media (Eliot, 2020 being a notable exception), and—consequently—politicians. It seems important, however, to foster a public discourse about the implications of the leeway car manufacturers have in responding to the lack of post-collision behavior specifications. An open question not empirically addressed so far is the importance people ascribe to this issue and whether or not car owners are willing to bear the costs of the required technology if AVs' equipped technology is not explicitly regulated. In addition, the interests of many vulnerable road users, such as pedestrians and cyclists, might be completely neglected if AVs' equipment is left entirely to the manufacturers' discretion.

The "data-driven study of driverless car ethics" enriches the academic ethical discourse by systematically incorporating ordinary people's moral intuitions and preferences (Bonnefon, Shariff, & Rahwan, 2016). Ultimately, they are who will be affected by AVs, either as passengers or as vulnerable road users, and they will determine how quickly AVs will prevail on the road (e.g., Adnan, Nordin, Bahruddin, & Ali, 2018; Awad et al., 2018; Bonnefon et al., 2016; Cunningham et al., 2019). The "moral machine" experiment is currently the largest study in this domain (Awad et al., 2018). Its authors analyzed millions of individuals' answers to modern variants of the so-called "trolley problem" (Foot, 1967; Greene, 2013). In this infamous dilemma, respondents choose whom a fully automated vehicle should run over in case of an unavoidable collision. In these dilemmas, people typically think AVs should save humans over animals, younger over older lives, and more over fewer lives (e.g., Awad et al., 2018; Bigman & Gray, 2020). For their own cars, however, people prefer AVs that save their passengers at all costs (Bonnefon et al., 2016). This points to a "social dilemma" (Bonnefon et al., 2016) where people find AVs with certain algorithms more moral but would prefer to own AVs with other algorithms. This is particularly relevant when regulations are either insufficient or absent.

Despite a substantial body of literature and a vast amount of data concerning preferences for pre-collision behavior, nothing is known about people's preferences for AVs' post-collision capabilities. Specifically, it is unclear whether people expect AVs to behave as human drivers if this behavior has no bearing on the accident's direct consequences. For instance, do certain behaviors—such as stopping at the accident site, calling the police, or helping determine responsibilities—have value in and of themselves? After all, a hit-and-run accident caused by a human driver is punished severely in general, not only if the victim(s) could have been saved. Thus, it could reasonably be equally unacceptable for AVs to continue driving after accidents, possibly because the AV did not register the accidents in individual cases (e.g., a large self-driving truck on interstate highways). How important would appropriate AV capabilities be to the public? Would people be willing to pay for them? What if the necessary technology for accident registration delays the introduction of AVs altogether?

If people value appropriate post-collision behavior, this valuation should be measurable from various angles. First, we investigated whether appropriate post-collision behavior enters the moral evaluation of the crash scenario. To elicit the moral value of post-collision behavior, we modified the trolley dilemma to explicitly entail information about the AV's behavior in an accident, something usually not included in standard trolley problems. In our modified trolley dilemma, one of the AV's two possible maneuvers led to appropriate post-collision behavior aligning with the 1968 Convention on Road Traffic. The other possible maneuver explicitly did not yield any of these behaviors due to AVs not detecting the accident in the first place. However, that post-collision behavior had no bearing on the accident's outcome in our study. Accidents were fatal no matter how the AV behaved in the aftermath. By manipulating which of the two possible maneuvers entailed the AV's appropriate post-collision behavior in our modified trolley problem, we attempted to infer the moral value of such behavior.

Second, we collected people's normative and empirical expectations about AVs' post-collision behavior by asking people what they think AVs are prepared to do and what AVs should do after accidents. Third, we elicited people's preferences for AVs with and without certain behavioral features regarding an accident and assessed people's willingness to pay for the required technology.

In our article's second section, we detail the materials and methods used in our study. In the third section, we present results. In the fourth section, we discuss the results, draw some conclusions, and emphasize future research questions on this topic.

Materials and methods

We conducted an online survey with 1,138 total participants between November 2020 and January 2021. All participants were residents of the United States recruited using Cloud-Research Prime Panels (Chandler, Rosenzweig, Moss, Robinson, & Litman, 2019; Litman, Robinson, & Abberbock, 2017). We chose Prime Panels because previous studies show that Prime Panels provide reliable survey results across many tasks and over many measures (Chandler et al., 2019). Moreover, Prime Panels' participants are required by default to pass screening questions before they can participate in a study (Chandler et al., 2019).

The survey was programmed via Qualtrics software and comprised four sections.[1] The first section contained general information about the survey, such as its approximate duration, contact information for the researchers responsible, and a query about voluntary participation in the study. Only participants who confirmed voluntary participation could complete the survey's other three sections.

The second section comprised a vignette study to assess the moral value of appropriate post-collision behavior, as defined by the 1968 Convention on Road Traffic. All participants were given a trolley-like scenario in which an AV could choose between two courses of action. In total, we used three scenarios in this section, with each participant exposed to only one scenario. Participants were randomly assigned to a specific scenario.

As the baseline scenario, we used a stochastic trolley problem from Krügel & Uhl (2020). We chose a stochastic trolley problem because it explicitly captures important risk management elements associated with AVs. Standard trolley problems are frequently criticized with respect to AVs because these problems disregard risks and are therefore unrealistic regarding real road traffic (Goodall, 2016b; Trussell, 2018; Winfield, Michael, Pitt, & Evers, 2019). Standard trolley problems rely on unlikely crash scenarios with little practical relevance for AVs (e.g., Himmelreich, 2018; Lundgren, 2020). For critics of trolley problems, the AVs' main task is assessing and distributing (minimal) accident risks in road traffic (e.g., Goodall, 2016a, 2016b), rather than solving ethical dilemmas in unavoidable accident scenarios (e.g., Trussell, 2018; Winfield et al., 2019). The main problem here is that moral judgments in the former might categorically differ from moral judgments in the latter (e.g., Fried, 2012; Nyholm & Smids, 2016).

Krügel & Uhl (2020) found no evidence of this categorical difference. In their study, moral judgments in trolley problems appeared independent regarding whether the

dilemma's situation is based on an inevitable accident scenario or on the distribution of marginal risks between road users. Nonetheless, we chose a stochastic trolley problem in our study to accommodate potential critics. In the specific scenario, an AV without passengers drives on a two-lane roadway and approaches a narrowing road. If the vehicle continues in its current lane, there is a 0.2% risk that the vehicle will fatally hit five road workers. If the vehicle changes lanes instead, there is a 1.0% risk of fatally hitting a single road worker on the other side of the road. In the study of Krügel & Uhl (2020), participants choices were halfway split between both actions. We wanted to allow for post-collision behavior to potentially influence the moral evaluation of the ethical dilemma in either direction, so we chose this as our baseline scenario. Therefore, a priori, the post-collision behavior's influence on the moral evaluation of the situation was possible to about the same extent in both directions.

As usual in trolley problems, our baseline scenario contained no information regarding the AV's behavior in case of an accident. Trolley problems elicit the influence of structural factors (from the setting or from the personal characteristics of potential victims) on the moral judgment regarding the underlying ethical dilemma. Post-collision behavior usually plays no role in trolley problems.

To assess whether post-collision behavior carries value on its own, we conducted two additional treatments in which we varied the AV's behavior regarding an accident. In both treatments, we used the baseline treatment scenario and added an additional paragraph describing the AV's post-collision behavior. If the AV could register an accident, it acted in accordance with Article 31 of the Convention on Road Traffic (UNECE, 2021). In this case, the vehicle (i) stopped at the accident site, (ii) called the police, and (iii) recorded the accident to later determine responsibilities. If, instead, the AV could not register an accident, none of these actions were initiated. No one else had observed the accident according to our scenario description, so, strictly speaking, a hit-and-run accident occurred. However, we did not explicitly mention a hit-and-run in our description to the participants because we feared the term itself might have negative connotations.

To observe post-collision behavior's potential influence on people's moral evaluation of the ethical dilemma, we introduced the ability to register accidents diametrically in the two treatments. In one treatment, the AV could register the accident with the five road workers but not the accident with the single road worker (*Hit-One-and-Run* in the following). The other treatment was just the opposite—the AV could register the accident with the single road worker but not the accident with five workers (*Hit-Five-and-Run* in the following). To isolate the importance of post-collision behavior itself, the registration of accidents

---

[1] Screenshots of the entire survey with all questions and treatments appear in the Online Supplementary Material.

did not affect the outcome for victims of the collision. If a collision occurred, the accident was fatal whether it was registered or not.

Participants were asked two questions regarding the respective scenario. First, they had to choose which of the two possible actions the AV should adopt: whether the AV should go straight ahead and take the 0.2% risk of a fatal collision with five road workers or instead change lanes and take the 1% risk of a fatal collision with a single road worker. After answering this question, participants were asked to indicate the relative morality of the two possible actions on a scale from zero to 100. Here, zero indicated going straight as morally better, 100 indicated changing lanes as morally better, and 50 indicated "in between."

Lastly, in this section participants had to answer a comprehension question about the scenario they witnessed. In vignette studies of this type, it is usually sufficient to ask relatively easy questions to obtain a test of the participants' attention. In similar studies, therefore, only between 10 and 20% of participants fail these questions (e.g., Bonnefon et al., 2016; Bigman & Gray, 2020; Krügel & Uhl, 2020). In the present study, participants crucially must understand the specific information regarding post-collision behavior. Therefore, we chose a relatively difficult comprehension question about the precise information regarding the AV's ability to register accidents in the scenario. Only about 50% of participants could answer this question correctly. In this way, we ensured the participants had read and understood the relevant information. We aimed for about 200 participants per treatment who could answer the comprehension question correctly, so we kept inviting new participants to the survey until reaching this goal. As a result, we ended up with 1,138 survey participants; of these, however, only 612 could answer the respective comprehension question correctly. As in similar studies, our data analysis is limited to those 612 participants who passed the comprehension question.

In the third section of the survey, we asked participants general questions about AVs' post-collision behavior. That is, we assessed participants' empirical and normative expectations about AVs' post-collision behavior. We measured their preferences for AVs with or without certain capabilities for post-collision behavior, and we asked about participants' willingness to pay for the required technology. In this section, all participants faced the same questions, regardless of the treatment to which they were previously assigned.

In the survey's fourth and final section, we gathered participants' demographic characteristics as well as their excitement and fear regarding a future with AVs and their willingness to travel in AVs as soon as they are available. Again, all participants received the same questions regardless of the treatment. After completing the fourth section, the survey was completed.

Results

The moral value of and expectations for AVs' post-collision behavior

In this section, we first examine whether appropriate post-collision behavior itself carries a moral component. That is, we investigated whether post-collision behavior influenced participants' moral evaluation of the underlying dilemma. Thereafter, we look at whether participants expected AVs to behave according to the guidelines of the Convention of Road Traffic after an accident and whether AVs should behave this way.

A total of 1,138 participants completed our study, but only 612 participants could answer the comprehension question about the described scenario correctly. As mentioned earlier, it was important for participants to properly read and understood the conditions of the AV's post-collision behavior in our scenarios. Therefore, we opted for a difficult comprehension question designed to check this. As is common in these types of studies (e.g., Bonnefon et al., 2016; Bigman & Gray 2020; Krügel & Uhl, 2020), our data analysis was based only on participants who passed the comprehension test. In our view, the strong selection of participants according to the comprehension question signals our data's quality.

Of those 612 participants, 215 (35%) were men and 560 (92%) had a driver's license. The average age was 44.85 years, and participants ranged from 18 to 87 years of age. Regarding participants' living environment, our sample had a full spectrum from rural to urban. Overall, participants were reluctant about a future with AVs. On a scale from 0 to 6, they were on average more fearful (3.89) than excited (2.65) about AVs, and their average willingness to travel with them was limited (2.45).

Randomly assigning participants to various treatments via Qualtrics worked reasonably well. Except for age, participants did not differ between treatments in any variables we collected. In the Hit-One-and-Run treatment, participants were slightly (but statistically significantly) younger than in the other two treatments (see Table in the Appendix).

Figure 1 depicts the moral value of AVs' appropriate post-collision behavior as judged by participants. The upper plot shows the proportion of participants who would instruct the AV to change lanes to spare the group of five and kill the single worker in each treatment. The lower plot shows participants' ratings of the relative morality of changing lanes versus continuing straight on. The baseline treatment contained no information about the AV's behavior regarding an accident, as is common in trolley-like dilemmas. In the Hit-One-and-Run treatment, the

Fig. 1 The moral costs of hit-and-runs. *Notes* The figure shows participants' responses to the vignettes in each treatment. **a** Shows the proportions of participants who would divert the vehicle towards the single worker to spare the five workers. **b** Shows the average ratings of the relative morality of diverting the vehicle compared to continuing straight. Boxes display the 95% confidence intervals. *P*-values are based on *chi-squared* tests in (a) and *t*-tests in (b)

(a) Changing lanes towards single worker

(b) Relative morality of changing lanes vs. continuing straight on
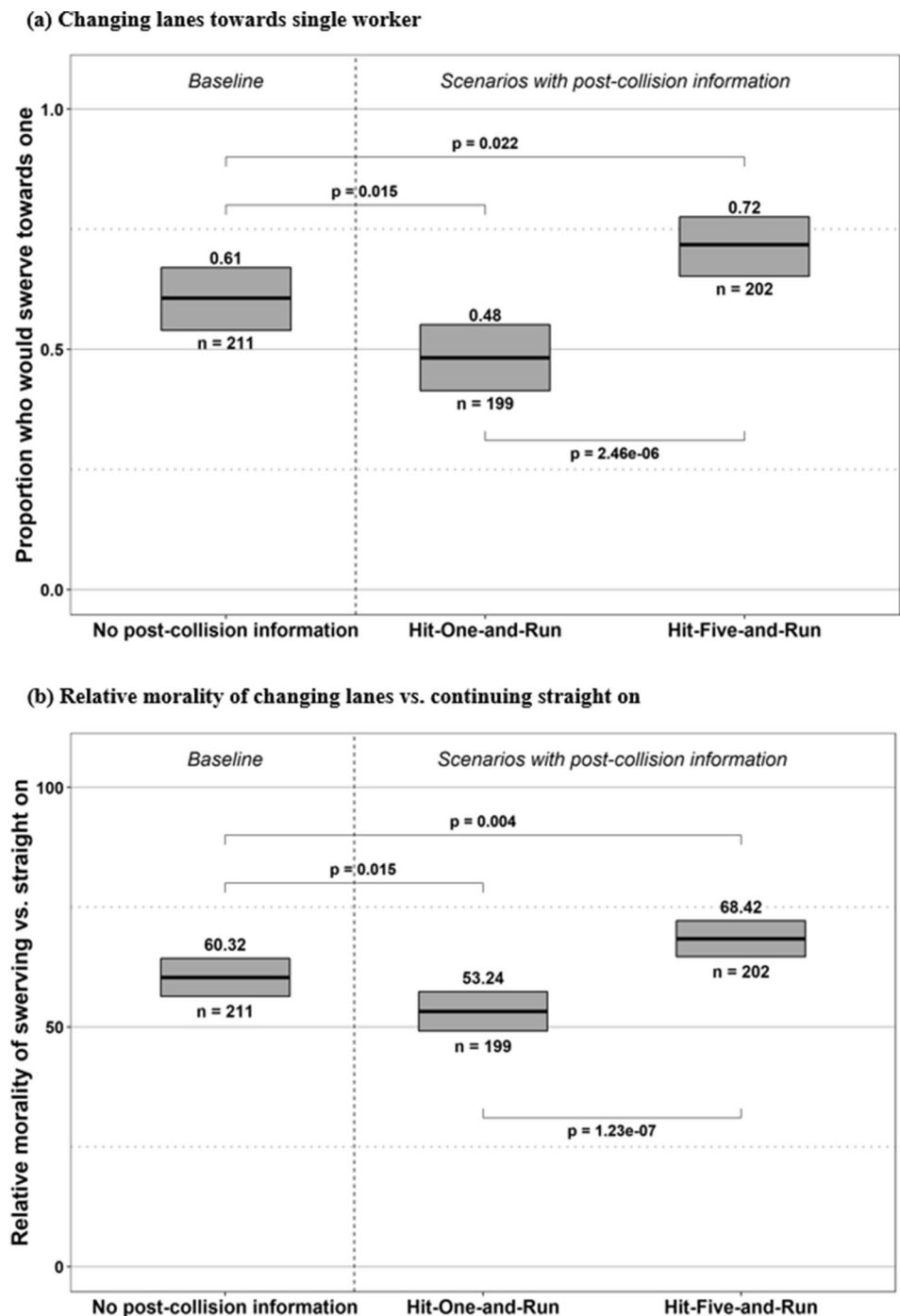
AV could register the accident with the five workers and behave accordingly, but it could not register the accident with the single worker. In the Hit-Five-and-Run treatment, this was the other way around.

Compared to 61% in our study and 50% in Krügel and Uhl's (2020) study, a slightly higher proportion of participants in our baseline treatment thought the AV should change lanes to the disadvantage of the single worker. Thus, despite that both maneuvers resulted in statistically comparable fatalities, participants in the present study slightly preferred avoiding less frequent but more severe accidents with many deaths. This was also reflected in participants' ratings of the relative morality of the two maneuvers in the baseline

Table 1 Regression analysis for decisions to swerve and relative morality of swerving

	Swerve (=1)			Relative morality of swerving		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	−0.226	−0.591**	−0.263	52.84***	44.39***	51.93***
	(0.273)	(0.273)	(0.276)	(5.60)	(5.52)	(5.86)
Hit-One-and-Run	−0.284**	−0.292**	−0.285**	−6.64**	−6.71**	−6.71**
	(0.128)	(0.129)	(0.128)	(2.91)	(2.90)	(2.92)
Hit-Five-and-Run	0.320**	0.340***	0.316**	8.09***	8.30***	7.96***
	(0.131)	(0.132)	(0.131)	(2.80)	(2.79)	(2.79)
Male (=1)	0.136	0.168	0.127	5.70**	5.82**	5.51**
	(0.115)	(0.114)	(0.116)	(2.46)	(2.45)	(2.49)
Age	0.005*	0.005	0.006*	0.07	0.07	0.08
	(0.003)	(0.003)	(0.003)	(0.06)	(0.06)	(0.07)
Excited about AVs	−0.034			−1.12*		
	(0.028)			(0.65)		
Fearful about AVs		0.092***			1.69**	
		(0.031)			(0.71)	
Travel in AVs			−0.021			−0.78
			(0.029)			(0.69)
Living in city	−0.006	−0.012	−0.008	0.47	0.34	0.44
	(0.028)	(0.028)	(0.028)	(0.63)	(0.62)	(0.63)
Driver's license (=1)	0.329*	0.293	0.321	3.74	2.99	3.48
	(0.199)	(0.202)	(0.198)	(4.16)	(4.20)	(4.14)
Observations	612	612	612	612	612	612

The table shows the coefficients and, in parentheses, robust standard errors of probit (Eqs. 1—3) and OLS (Eqs. 4—6) regressions, respectively. In regressions (1) through (3), the dependent variable is the participants' decision whether the AV should change lanes to the detriment of the single worker (Yes: 1; No: 0). In regressions (4) to (6), the dependent variable is the participants' rating of the relative morality of changing lanes compared to continuing straight on (between 0 and 100)

$*p < 0.1 \ **p < 0.05 \ ***p < 0.01$

treatment. With an average rating of 60.32, changing lanes was perceived as more moral compared to going straight.[2]

Adding appropriate post-collision behavior to one of the two maneuvers shows that post-collision behavior has moral value in and of itself. If the AV registered an accident involving five workers and behaved appropriately but did not register an accident involving the single worker, significantly fewer participants thought the AV should change lanes to shift the accident risk to the single worker. In contrast, if the AV registered an accident with a single worker but not with a group of five, significantly more participants (compared to the baseline treatment) thought the AV should change lanes. To identify whether appropriate post-collision behavior has moral value in itself, the two treatments with information

on post-collision behavior must be compared. Relative to the Hit-One-and-Run treatment, 50% more participants ($= \frac{0.72}{0.48} * 100$) selected changing lanes, to the detriment of the single worker, as the appropriate maneuver if the accident with the single worker (instead of the five workers) was registered by the AV (i.e., Hit-Five-and-Run). Considering that registering the accident changed nothing for our scenarios' victims, the difference between the two treatments appears substantial. AVs' ability to behave appropriately after an accident has apparent spillover effects on the moral evaluation of the dilemma prior to an accident. This is also reflected in the ratings of the relative morality of the two maneuvers in the corresponding treatments (see Fig. 1[b]).

Probit and OLS regressions with various control variables confirm the results in Fig. 1. Equations (1) through (3) in Table 1 present the results of probit regressions, with the decision of whether to change lanes to the detriment of the single worker as the dependent variable. Equations (4) through (6) present the OLS regressions' results, with the relative morality of changing lanes versus going straight as the dependent variable. Independent variables in each case were all other variables collected in addition to treatment

---

[2] Subjects in this study slightly preferred avoiding less frequent but more severe accidents, suggesting the moral relevance of decomposing an accident risk into its constituents: probability and magnitude. Thus, applying stochastic trolley dilemmas could provide a useful tool for a societal debate on the risk ethics of AVs that focuses on more realistic pre-collision behavior that arises from a stochastic environment.
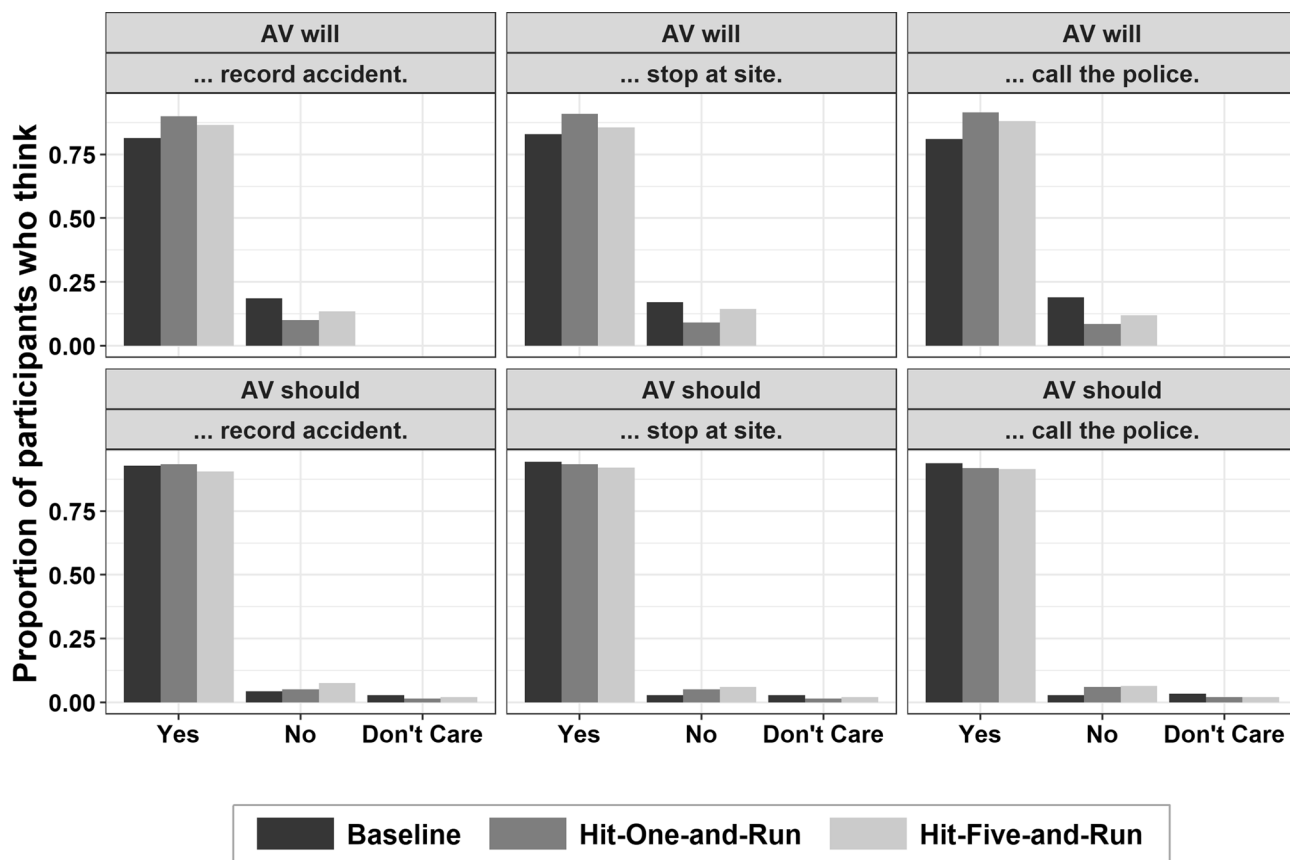
Fig. 2 Empirical and normative expectations regarding AVs' post-collision behavior. The top row of the figure shows the participants' empirical expectations regarding AVs' post-collision behavior, and the bottom row shows the participants' normative expectations. Ques-tions regarding normative expectations contained a third response category "*I don't care*," which did not exist for the questions regarding empirical expectations

dummies. The variables *Excited about AVs*, *Fearful about AVs*, and *Travel in AVs* strongly correlated with each other, so we included each of these three variables separately in the regressions.

The regressions confirm post-collision behavior's significant effect depicted in Fig. 1. If the AV registered the accident with the five workers but not the accident with one worker, participants' willingness to change lanes to the detriment of the single worker and the relative morality of changing lanes decreased significantly compared to the baseline treatment. In contrast, if the AV registered the accident with the single worker but not with the five workers, participants' willingness to change lanes and the relative morality of changing lanes increased significantly. In addition, men rated the relative morality of changing lanes slightly higher than did women, and participants who had more fear of AVs overall were more inclined to change lanes in our scenarios. Although speculative, fearful participants may focus more on the collision's consequences in lieu of accident probabilities and therefore tend towards saving more lives rather than fewer. A useful analogy might be a fearful flight passenger

obsessed with the event of a crash despite its extremely low probability.

The lower row in Fig. 2 clearly shows that participants believe AVs should be capable of appropriate post-collision behavior as defined by the Convention of Road Traffic. Irrespective of the treatment, 91% to 94% of participants said AVs should (i) call the police in case of an accident, (ii) stop at the accident site, and (iii) record the accident to determine responsibilities. Interestingly, a substantial majority of participants (between 81 and 91%) also thought AVs would possess these capabilities. Apparently, participants did not know that AVs currently do not fully have, and may not be required to have, these capabilities.

Participants' preferences and their willingness to pay for the required technology

In addition to the empirical (i.e., "what will be?") and normative (i.e., "what should be?") expectations regarding AVs' post-collision behaviors, we investigated the value of this behavior through participants' purchasing preferences.

Fig. 3 Preferences for AVs with or without post-collision capabilities. Notes*: The figure shows participants' average preferences for AVs with and without capabilities for corresponding post-collision behaviors. Participants could express their preferences on a slider between 0 and 100, with 50 labeled as "*in between*" in each case. Error bars display standard errors of the mean
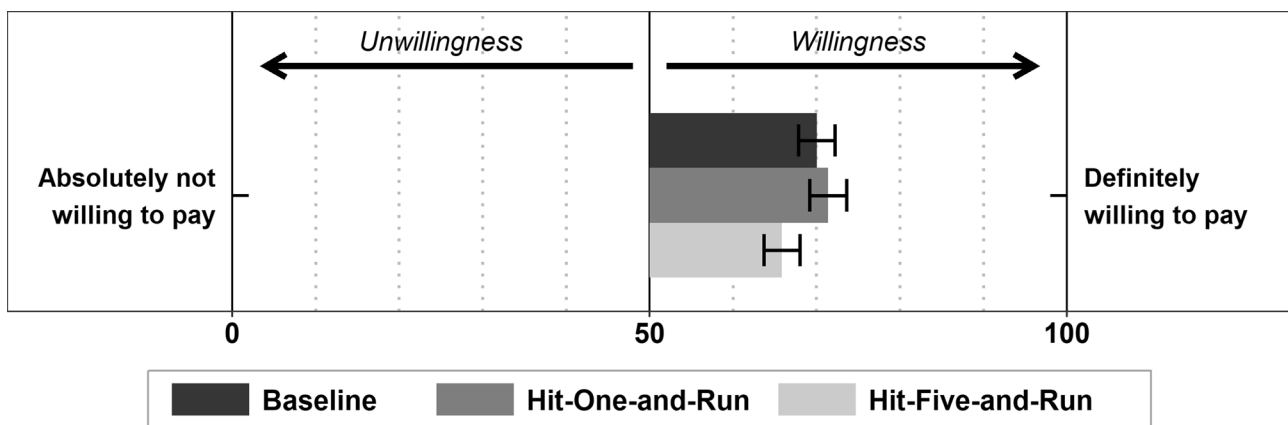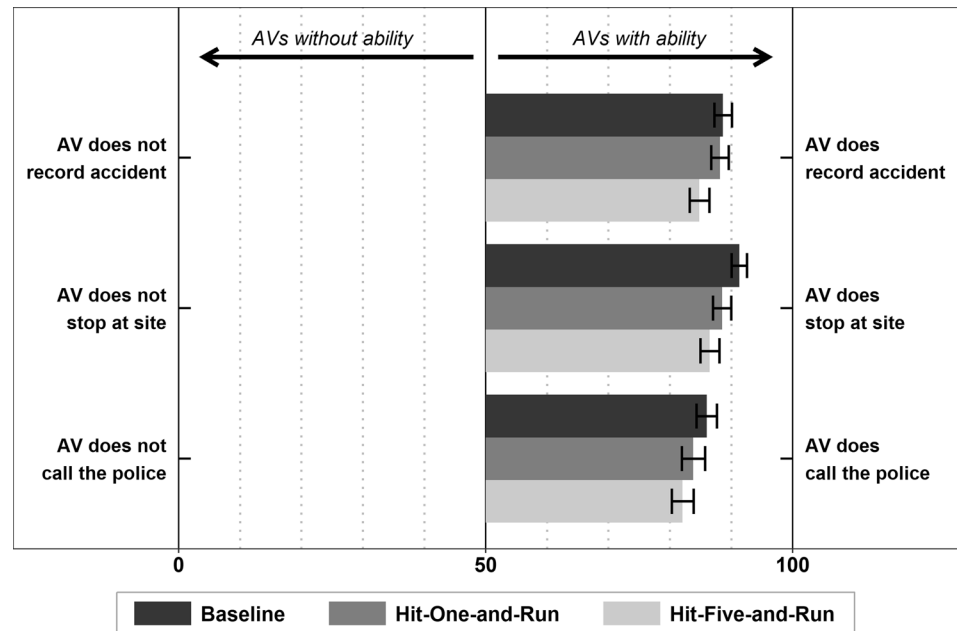

Fig. 4 Willingness to pay for devices enabling appropriate post-collision behavior. Notes: The figure shows participants' average stated (un-)willingness to pay to provide AVs with the necessary equipment to be capable of appropriate post-collision behavior. Participants could express their (un-)willingness to pay for these devices on a slider between 0 and 100, with 50 labeled as "in between." Error bars display standard errors of the mean

For each of the three possible behaviors of (i) recording the accident, (ii) stopping at the accident site, and (iii) calling the police, we asked participants whether they would more likely buy an AV with or without these capabilities. Here, participants were asked to imagine there were no legal regulations regarding AVs' post-collision behavior and that the corresponding cars would not differ in quality or in price. Participants could indicate their preferences via a slider between 0 and 100, where "0" indicated preference for an AV without a capability, "100" indicated preference for an AV with the respective capability, and 50 indicated "in between."

As Fig. 3 shows, participants expressed a strong preference for AVs with capabilities for all three post-collision behaviors (with an average rating between 82 and 91). Thus, participants stated that AVs should have these capabilities and that they would choose these types of AVs. Figure 4 suggests the participants would even pay for the AVs' necessary equipment. Of course, these statements are purely hypothetical and abstract. However, with an average rating between 65 and 71, participants did not indicate any reluctance to bear the costs of the required equipment.

Table 2 shows the results of OLS regressions exploring whether participants' preferences (Eqs. 1–3) and willingness to pay (Eqs. 4–6) associated with our demographic

Table 2 Regression analysis for preferences and WTP for post-collision devices

	Preferences for devices			Willingness to pay for devices		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	75.31***	78.93***	74.67***	46.86***	64.89***	47.53***
	(3.86)	(3.70)	(3.91)	(6.33)	(6.52)	(6.53)
Hit-One-and-Run	−2.94	−2.98	−2.82	−3.42	−3.45	−3.10
	(1.82)	(1.83)	(1.82)	(3.01)	(3.11)	(3.06)
Hit-Five-and-Run	−1.94	−1.81	−1.84	0.60	0.72	1.02
	(1.78)	(1.80)	(1.78)	(3.06)	(3.11)	(3.07)
Male (= 1)	−4.46**	−3.66**	−4.70***	−7.47***	−5.52**	−7.51***
	(1.74)	(1.73)	(1.76)	(2.72)	(2.78)	(2.82)
Age	0.27***	0.24***	0.28***	0.23***	0.16**	0.23***
	(0.04)	(0.04)	(0.05)	(0.07)	(0.07)	(0.07)
Excited about AVs	1.00***			3.74***		
	(0.38)			(0.69)		
Fearful about AVs		−0.10			−1.98**	
		(0.44)			(0.78)	
Travel in AVs			1.09***			3.26***
			(0.39)			(0.74)
Living in City	0.87**	0.94**	0.88**	1.78***	2.10***	1.86***
	(0.41)	(0.42)	(0.41)	(0.67)	(0.68)	(0.68)
Driver's License (= 1)	−3.33	−3.12	−3.07	−0.45	0.85	0.46
	(2.57)	(2.56)	(2.61)	(4.69)	(4.95)	(4.86)
Observations	612	612	612	612	612	612

Notes: The table shows the coefficients and, in parentheses, robust standard errors of OLS regressions. In regressions (1) through (3), the dependent variable is the participants' mean preference for AVs with the abilities to (1) record accidents, (2) stop at the accident site, and (3) call the police. In regressions (4) through (6), the dependent variable is the participants' stated willingness to pay for the necessary technology in AVs. In each case, the dependent variables vary between 0 and 100

*$p < 0.1$ **$p < 0.05$ ***$p < 0.01$

variables. Because participants' preferences for each capability highly correlated with each other, we calculated the average preference for AVs with appropriate post-collision behaviors based on all three capabilities.

As seen in Table 2, participants' preferences and their willingness to pay for required devices were independent of the treatment they were assigned to. Compared to women, men's preferences for AVs with appropriate post-collision behavior appeared weaker, and they were less willing to pay for the necessary equipment. In addition, older participants showed a greater preference and willingness to pay for AVs with appropriate post-collision behaviors. Likewise, participants in urban environments valued these AV features more than did participants in rural areas. Finally, participants who were more excited about AVs and more willing to travel with them were also more likely to prefer AVs with appropriate post-collision behaviors, and they were more willing to pay for necessary devices. In contrast, participants fearful of AVs were less willing to pay for the technical equipment for post-collision behaviors.

Conclusion

Our study's results suggest that people care strongly about AVs' post-collision behavior. Their respective preference is so strong that they even adapt their decisions regarding appropriate pre-collision behavior to whether the AV can detect the accident and could thus initiate appropriate post-collision behaviors. People's clear preference for the pre-collision behavior that enables the appropriate post-collision behavior is supported by their evaluation that it has a higher morality relative to the alternative behavior.

Furthermore, people have strong empirical expectations that the AV will call the police, record the accident, and stop at the site. This is particularly noteworthy considering the current reality of a regulatory gap concerning

AVs' post-collision behavior. People's empirical beliefs that AVs will engage in certain post-collision behaviors (that are not actually provided) are backed by their normative preferences that decisively express the opinion that this should be the case. They even express a substantial, though abstract, willingness to pay for the technological features that enable such post-collision behaviors.

The finding that people attach significant value to AVs' post-collision behavior underlines the importance of detecting accidents as completely as possible. However, this might imply that accidents occurring as an unintended result of certain maneuvers are less likely to be detected. Changing lanes, for instance, naturally increases the probability of lateral as opposed to frontal collisions. In turn, lateral collisions might be less likely to be detected due to weaker sensor technology on the vehicle's side. Taking people's ethical preferences seriously would require their intuitions be adequately reflected in the AV's pre-collision decisions. Specifically, the AV's algorithm must penalize maneuvers with a higher risk of failing to detect an accident.

Our studies have several limitations. First, the preferences elicited in the experiments are based on a sample of US residents. Participants from different cultural backgrounds may express different ethical views regarding the importance of post-collision behavior. A systematic cross-cultural comparison of ethical intuitions therefore seems worthwhile. Second, it is unclear whether people place an instrumen-

the inherent value of post-collision behavior by mentioning the recording of the accident solely in relation to determining responsibilities. This determination of responsibilities and knowledge about events might have a general value in and of itself. However, we cannot exclude the possibility that participants believed that recording the accident also increases road safety and consequently reduces the risk of fatal consequences. For instance, participants might believe the accident report will help close safety gaps by systematically evaluating the accident data, making autonomous driving safer in the future.

This article might constitute a starting point for deeper inquiries into people's ethical preferences on AVs' post-collision behavior. Given the topic's apparent importance for the affected, failing to address the issue would constitute a serious omission from an ethics perspective. The questions that remain certainly yield arguments for extending the field of data-driven ethics of autonomous cars to include a research program on post-collision behavior.

Appendix

See Table 3.

Table 3 Demographic characteristics of participants

	Baseline trolley	Trolley with post-collision information		Treatment differences
		Hit-One-and-Run	Hit-Five-and-Run	
Participants	211	199	202	
Male	33%	34%	38%	$p = 0.548$
Driver's License	92%	91%	91%	$p = 0.857$
Age	46.73	41.54	46.15	$p = 0.009$
	(18.97)	(16.74)	(19.62)	
Living environment	3.19	3.26	3.41	$p = 0.482$
(0:rural … 6:city)	(1.97)	(1.89)	(1.84)	
Excited about AVs	2.53	2.67	2.74	$p = 0.555$
(0:not at all … 6:for sure)	(2.14)	(2.00)	(1.89)	
Fearful about AVs	4.00	3.94	3.74	$p = 0.336$
(0:not at all … 6:for sure)	(1.87)	(1.73)	(1.85)	
Willing to travel in AVs	2.4	2.46	2.51	$p = 0.848$
(0:not at all … 6:for sure)	(2.05)	(1.97)	(1.9)	

The table shows the demographic characteristics of all participants in each experimental treatment who answered the control question correctly. For *Age* and all subsequent characteristics, the mean values and, in parentheses, the standard deviations are reported. In the column "Treatment differences," the *p*-values of corresponding statistical tests are provided, where it was tested whether the treatments differed in any of the characteristics. For *Male* and *Driver's License*, the *p*-values are based on *2 × 3-Fisher's Exact tests* and for all other characteristics on *one-way ANOVA tests*

tal or inherent value on certain post-collision behaviors of AVs. With our scenario descriptions, we tried to address

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval German Association for Experimental Economic Research.

Consent to participate Written consent was obtained, see supplementary material.

References

AAA Foundation for Traffic Safety. (2018). *Hit-and-run crashes: Prevalence, contributing factors and countermeasures*. Research Brief. https://aaafoundation.org/wp-content/uploads/2018/04/18-0058_Hit-and-Run-Brief_FINALv2.pdf

Adnan, N., Nordin, S. M., Bahruddin, M. A. B., & Ali, M. (2018). How trust can drive forward the user acceptance to the technology? In-vehicle technology for autonomous vehicle. *Transportation Research Part A: Policy and Practice, 118*, 819–836.

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

Bigman, Y. E., & Gray, K. (2020). Life and death decisions of autonomous vehicles. *Nature, 579*(7797), E1–E2.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573–1576.

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods, 51*(5), 2022–2038.

Cunningham, M. L., Regan, M. A., Horberry, T., Weeratunga, K., & Dixit, V. (2019). Public opinion about automated vehicles in Australia: Results from a large-scale national survey. *Transportation Research Part a: Policy and Practice, 129*, 1–18.

Economic Commission for Europe. (2020). *Revised Amendment proposal to the 1968 Convention on Road Traffic*. Economic Commission for Europe, Inland Transport Committee, Global Forum for Road Traffic Safety. https://unece.org/fileadmin/DAM/trans/doc/2020/wp1/ECE-TRANS-WP1-2020-1-_Rev1e_.pdf

Eliot, L. (2020). *On the hit-and-maybe-run self-driving car dilemma (akin to 'The Molly Problem')*. Forbes October 14. https://www.forbes.com/sites/lanceeliot/2020/10/14/on-the-hit-and-maybe-run-self-driving-car-dilemma-akin-to-the-molly-problem/?sh=561d183a4662

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5*, 1–7.

Fried, B. H. (2012). What does matter? The case for killing the trolley problem (or letting it die). *The Philosophical Quarterly, 62*(248), 505–529.

Goodall, N. J. (2014). Machine ethics and automated vehicles. *Road vehicle automation* (pp. 93–102). Springer.

Goodall, N. J. (2016a). Can you program ethics into a self-driving car? *IEEE Spectrum, 53*(6), 28–58.

Goodall, N. J. (2016b). Away from trolley problems and toward risk management. *Applied Artificial Intelligence, 30*(8), 810–821.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Atlantic Books.

Himmelreich, J. (2018). Never mind the trolley: The ethics of autonomous vehicles in mundane situations. *Ethical Theory and Moral Practice, 21*(3), 669–684.

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.

Krügel, S. & Uhl, M. (2020). Autonomous Vehicles and Moral Judgments under Risk. Available at SSRN 3686613.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442.

Lundgren, B. (2020). Safety requirements vs. crashing ethically: What matters most for policies on autonomous vehicles. *AI & SOCIETY*, 1–11.

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.

Trussell, H. J. (2018). Why a special issue on machine ethics. *Proceedings of the IEEE, 106*(10), 1774–1776.

UNECE. (2021). *Road traffic and road signs and signals agreements and conventions*. https://unece.org/road-traffic-and-road-signs-and-signals-agreements-and-conventions

Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems. *Proceedings of the IEEE, 107*(3), 509–517.

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