The Role of Utilitarianism, Self-Safety, and Technology in the Acceptance of Self-driving Cars

Stamatis Karnouskos

Abstract Self-driving vehicles are currently being developed in major industries for mass public deployment. Their benefits are attractive, and there is interest in having them on the streets as quickly as possible. Delegating responsibility to the cars has far-reaching impacts. As accidents will be unavoidable, the self-driving car will be asked to make life and death decisions that will impact human lives. However, the impact of machine decision making on the overall acceptance of self-driving cars in society is far from sufficiently addressed. The research presented here investigates three factors, i.e., Technology, Self-Safety, and Utilitarianism, and hypothesizes their link to self-driving car acceptance. Survey-collected data is statistically analyzed in order to assess the proposed hypotheses. The empirical outcome of this work shows that the three investigated factors contribute to the societal acceptance of self-driving cars, with technology being the major contributor while the ethical aspects (self-safety and utilitarianism) follow.

1 Introduction

Advanced and autonomous vehicles are increasingly becoming a reality worldwide (Fernandez-Rojas et al, 2019), however, we are not ready for their mass introduction (Shladover and Nowakowski, 2019; Mordue et al, 2020). This delegation of responsibilities to sophisticated machines enables us to undertake complex actions, with little human effort, as the underlying complexity is hidden by the machines, e.g., in driver-assistance features of modern cars. Several areas utilize such intelligent machines that make mission-critical decisions autonomously,

S. Karnouskos
SAP, Walldorf, Germany
E-mail: karnouskos@ieee.org

and although most of them were up to recently in industrial settings, nowadays they increasingly penetrate areas that directly affect the general population. One key domain that is intensively experimenting with intelligent machines is the automotive sector, where traditional manufacturers and technology companies experiment with various levels of autonomy, e.g., Tesla (cars, trucks), Uber (cars), and Waymo (taxi service).

Modern cars already provide significant automated features to their drivers e.g., cruise control, parking, road deviation alarms, object detection, crash avoidance. However, while the awareness of the car itself and the warnings it can propagate to its driver increase due to the utilization of more sophisticated technology, we have to be pragmatic and reconcile with the view that not all of the accidents will be possible to be avoided. On the one hand, self-driving cars, for a variety of reasons, have already been involved in fatal accidents either of the driver or pedestrians (NTSB, 2019a,b). On the other hand, cars with self-driving capability have also been reported to have prevented serious pedestrian injuries, as well as driving its owner to the hospital during a life-threatening emergency (BBC, 2016).

Self-driving car technology promises benefits such as traffic efficiency, pollution reduction, and elimination of human-error related accidents (Ethik-Kommission, 2017). While the public may be positively inclined towards autonomous cars (Rödel et al, 2014), this attitude may vary depending on the level of automation and how this is offered (Kyriakidis et al, 2015). The car's decision-making capabilities are expected to improve in the future, as with multiple sensors the car will be able to acquire detailed situational awareness (Hussain and Zeadally, 2019; Karnouskos and Kerschbaum, 2018), which coupled with artificial intelligence may enable self-driving cars to be capable of anticipating and reacting to the environment better than humans (sim-

ple cases are already evidenced by modern driver assistance features of cars). However, such cars, although they are expected to come with significant benefits, also raise concerns (Ethik-Kommission, 2017; Li et al, 2019; IEEE, 2018; Hicks and Simmons, 2019; Bremner et al, 2019; Karnouskos, 2020; Bonnefon et al, 2016).

Despite the advances in algorithms and hardware, complex environments still lead to unexpected and erroneous behaviors on behalf of self-driving cars (Guo et al, 2019). In the context of imminent unavoidable accidents, self-driving cars will have to make timely decisions, based on their internal algorithms, on the actions they will follow to avoid accident or minimize its impact. This, in several cases, may translate to life and death decisions (Carsten et al, 2015; Karnouskos, 2020). Recent fatal crashes of self-driving cars have sparked interest in this area, as it is no longer an academic debate, but a reality. Nevertheless, it is still approached in an asymmetric way, mostly from a technology viewpoint, and to a lesser degree from a legal and ethical standpoint (Karnouskos, 2020). However, there is little towards empirical research that shows how such factors affect the social acceptance of self-driving cars.

This work devotes itself towards investigating the acceptance of self-driving cars via three factors i.e., utilitarianism, self-safety, and technology. From the ethical side, the focus is limited to the dilemmas and considerations captured in utilitarianism and self-safety angles. A deep dive specific to ethics is carried out in complementary work by Karnouskos (2020). The ethical dimensions are investigated together with the additional factor of technology, and to our knowledge, this combination has not been addressed empirically. Hence, it is hypothesized that utilitarianism, self-safety, and technology have an impact on the acceptance of self-driving cars. Via a survey with a population of ($n = 62$), quantified data is collected in order to evaluate the hypotheses. The contribution is the theoretical model linking the selected three factors to self-driving car acceptance, as well as its empirical assessment and validation.

The paper is structured as follows: after the introduction in section 1, a discussion on the key issues pertaining to self-driving car acceptance is discussed in section 2. The empirical data and the statistical analysis is carried out in section 3, while a critical discourse follows in section 4. Finally, the conclusions are laid out in section 5.

2 Self-driving Car Acceptance

Acceptance of self-driving cars may influence the success or failure of their public introduction (Hevelke and Nida-Rümelin, 2014; Lin, 2015; Rödel et al, 2014; Nees,

2016), and therefore it is worth investigating. Contemporary research is devoted mostly from a qualitative viewpoint to the benefits or challenges they bring or pose. Empirical research exists (Nees, 2016; Hohenberger et al, 2016; PWC, 2016; Bansal et al, 2016; Bansal and Kockelman, 2016; Haboucha et al, 2017; Zmud et al, 2016), where ethical aspects are either partly considered or are related to privacy (Karnouskos and Kerschbaum, 2018). Only recently, more explicit focus on the ethics of self-driving cars and their impact on their acceptance (Bonnefon et al, 2016; Hevelke and Nida-Rümelin, 2014; Rödel et al, 2014; Coca-Vila, 2017; Karnouskos, 2020; de Sio, 2017; Rhim et al, 2020), is emerging.

Technology plays a role in how people adopt innovations, and the same holds for self-driving cars. Hypothetical situations involving ethics and risk are not new, and for instance, the "trolley" dilemma is well-known in experimental philosophy, although some voice the opinion that such dilemmas are policy and engineering distractions in the real world scenarios (Freitas et al, 2019). The question approached here is that of mixing ethics and technology, in the context of unavoidable accidents. An interesting question is raised in this context, which is what people think about self-driving cars and their autonomous decisions when these "intelligent machines" have to make real-world decisions impacting the well-being of other humans, e.g., life and death situations in civilian context (we explicitly distance ourselves from the military usage of autonomous vehicles and weapons).

From the ethics side, utilitarianism considers as the best action the one that produces the most good. Hence, in the hypothetical scenario of a critical situation where the choice lies between that of sacrificing five pedestrians or the two car passengers, the self-driving car decision would probably be (assuming the number of lives lost is the only criterion) to kill the car passengers (instead of the pedestrians). On the opposite side, one might consider the "self-safety" first approach, which implies that self-driving car passengers are primarily protected by the car, and everything else is a best-effort which would imply saving the passenger lives at all cost even if that means the fatal injury of the five pedestrians.

Utilitarianism and self-safety were selected because they exemplify such a dilemma in the self-driving car decision-making process and the challenges it brings. There is, however, no clear position on what should be done and its implications, e.g., if utilitarian ethics dictate the behavior of the car, people may not buy them (Malle et al, 2015), as their own car would harm them to save strangers. This effectively will limit the purchases of such cars, and their public introduction, which in

turn will limit the expected societal benefits (e.g., accident reduction). Therefore, more research is needed to understand the potential directions and their implications. Pertinent research with respect to unavoidable collisions shows that the morality of human decisions may indeed be a utility function linked to the value of life (Sütfeld et al, 2017).

Car acceptance is linked to ethics and technology, and several questions of interest are raised at large that are challenging (Mordue et al, 2020). What ethical framework should self-driving cars decision making be based upon? There are several that could be utilized e.g., as shown by Karnouskos (2020). Would ethical framework diversity be the norm, where individual citizens choose themselves what ethical decisions the car may make, or would the selection of ethics in the self-driving car be imposed by regulators, as all cars should make the same decisions in similar situations? Would technology and ethics correlate, e.g., could cars with better technology take better/more-timely ethical decisions? How would these considerations affect self-driving car acceptance? All of these are pertinent questions, and while this research touches on some of their issues, they should be investigated in a more detailed way. Nevertheless, they do set well the overall context for some of the insights and results presented here.

While self-driving cars are still the exception, experimentation is ongoing in order to reach a sophisticated level that could signal their public introduction at mass. A variety of stakeholders (Borenstein et al, 2017; Mordue et al, 2020) are implicated, e.g., technology companies, car manufacturers, legislators, user organizations, engineers, designers, etc. A dialogue that addresses decision making in critical situations needs to be properly addressed, and the role of ethics and technology, as well as their interrelationships and implications, need to be well understood. Some proposals/considerations on how to deal with such aspects exist, e.g., in Germany, a proposal was made (Ethik-Kommission, 2017) on the freedom to decide in conflict/dilemma situations (e.g., unavoidable accidents), the principle of minimizing damage but without putting a price on human life, etc.

3 Empirical Results

Acceptance of self-driving cars is seen as the major issue for their introduction in future business or civilian contexts. To address this issue, and in line with some existing surveys (Bonnefon et al, 2016; Kyriakidis et al, 2015; Karnouskos, 2020), this approach hypothesized that the three identified factors might impact the acceptance of self-driving cars, and more specifically, the three hypotheses (H1–H3) are: Technology (H1), Self-safety prioritization (H2), and Utilitarianism (H3) have an effect on self-driving car acceptance. There are different methods for obtaining the necessary data, e.g., interviews, group discussions, questionnaires, observation, and document studies. The questionnaires pose a good fit, as they can include questions and predefined answers which the respondent can answer, and is ideal for quantitative data collection (Johannesson and Perjons, 2012). Informed consent was presented, where all major issues are analyzed, and its (electronic) signing was a precondition to continue with the survey.

The empirical data acquired via an online survey is ordinal data described on a Likert scale, and all data were entered and semantically validated in real-time electronically. Overall there have been $N = 62$ responses. From a gender demographic point of view, 30.6% were females and 69.4% males. With respect to age, the majority, i.e., 43 of respondents are in the group $18 - 29$, 14 were in $30 - 44$ years old, and 5 were over 45 years old group. While this survey has a limited number of participants ($n = 62$), the sample is statistically sufficient as the different metrics show, to induce the respective correlations. Other empirical studies in the field also have similar participant numbers e.g., $n = 70$ (Rhim et al, 2020). The hypothesized factors were approached via a number of questions per factor, which were coded respectively: (T)echnology (T1,T2,T3), (S)elf-Safety (S1, S2, S3, S4), (U)talitarian (U1, U2, U3, U4), and self-driving car (A)cceptance (A1, A2, A3). Since all the variables are on the Likert scale, variables can be excluded only if they show no variance. Hence the primary focus is on kurtosis where values > 1 or < -1 may be problematic. Such values exist in the dataset for U1, U3, A1.

For the analysis, we have used typical empirical method indicators follow the method and steps, as discussed in (Karnouskos, 2020). The Kaiser-Meyer-Olkin (KMO) statistic is a Measure of Sampling Adequacy (MSA) and is calculated to be .747, which is characterized as middling but adequate. KMO, in conjunction with Bartlett's test of sphericity, that shows χ^2 is 48.877, with 91 degrees of freedom (DF) indicate that the dataset exploratory factor analysis (EFA) can be meaningfully carried out.

EFA was conducted with maximum likelihood, which is also used in the Structural Equation Modeling (SEM) in the IBM AMOS tool, and Promax because it can account for the correlated factors. Four factors are identified with eigenvalues greater than 1, which are responsible for a cumulative variance of 66.99%. The four factors that emerged from the EFA are in line with the initial four hypothesized ones considered from the theo-

retical framework, and therefore, no additional insights are evident.

Cronbach's α is a reliability measure and was carried out for all extracted factors separately i.e. Technology is above .7, which is characterized as acceptable, Utilitarianism and Acceptance are above .8, which is characterized as good, and Self-safety is above .9, which is characterized as excellent.

Structural Equation Modeling (SEM) was carried out, and in the process, several metrics are calculated: χ^2 (CMIN) is 84.552 and the more commonly used relative χ^2 (CMIN/DF) is 1.174. A CMIN/DF value < 2, as in this case, constitutes an acceptable fit. The Goodness of Fit Index (GFI) is .838, and the adjusted GFI (AGFI) is .764. Both of GFI and AGFI are < 1 (perfect fit) but near to it, which indicates an acceptable fit. The Comparative Fit Index (CFI) is .972, and as it is close to 1 (perfect fit) it is acceptable. The Root Mean Square Error of Approximation (RMSEA) is .053, which shows a good fit.
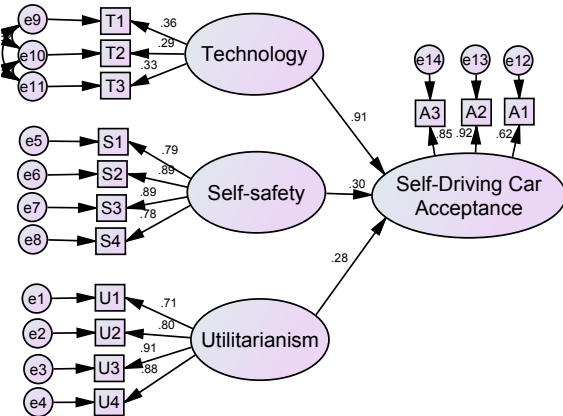

Fig. 1 Structural Equation Model in AMOS (with standardized estimates)

The hypothesized model has been constructed and executed in the IBM AMOS tool. AMOS enables the design of the model that shows the hypothesized relationships among variables, and its execution. All considered factors are shown as ovals in Figure 1, while the values imprinted on the arrows reflect path coefficients (standardized estimates), which show the weight of the links in the path analysis.

Table 1 Results of hypotheses testing

Hypothesis	Path	Path Coeff. Weight	CR value >1.96	Support Decision
H1	Technology → Acceptance	.913	2.501	Supported
H2	Self–Safety → Acceptance	.296	2.197	Supported
H3	Utilitarianism → Acceptance	.280	2.070	Supported

As a result, the interest is focused on the path coefficient weight from each factor towards the self-driving car acceptance entity. The path coefficient weight is also summarized in Table 1, including the calculation of the Critical Ratio (CR) metric. A CR less than -1.96 or greater than 1.96 indicates two-sided significance at the customary 5% level. The result derived from the SEM analysis and summarized in Table 1, shows that statistically, all three initial hypotheses posed are supported and hold for the specific dataset (empirically collected data). As such, the empirical data confirm that there is indeed a statistically significant link among the hypothesized factors, i.e., Technology (H1), Self-Safety (H2), Utilitarianism (H3) and that they impact Self-driving car acceptance.

4 Discussion

The aim was to quantitatively investigate if the three hypothesized factors, i.e., Technology, Self-Safety, and Utilitarianism, impact the self-driving car acceptance. Based on this, collected survey data, and after the rigorous statistical analysis with EFA and SEM the result was that: (i) the originally hypothesized model is plausible and represents a relatively good fit to the empirically measured data, and more importantly (ii) there is a strong indication for the link between the three identified factors and the user acceptance of self-driving cars.
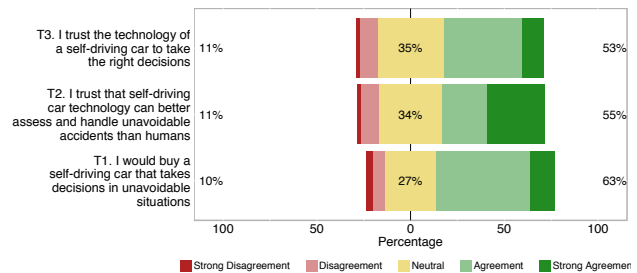

Fig. 2 Survey: Technology

It was hypothesized (H1) that technology behind self-driving cars may impact their acceptance. Technology is the key element that acts as a differentiator of the self-driving cars and the benefits they bring. People often cite the benefits of technology as a significant aspect when they consider an acquisition of a car, something that eventually got also technology companies interested in self-driving cars. This has been a pertaining issue in other surveys (Bonnefon et al, 2016; Kyriakidis et al, 2015; Sütfeld et al, 2017), and this research also confirm that there is a significant influence when peo-

4

ple decide to accept self-driving cars. By examining the questions posed in the survey for this factor (shown in Figure 2), one can see that the majority of people trust the technology of self-driving cars to make the right decisions. Over half of respondents also consider that in the case of unavoidable accidents, the car itself can most probably make a better decision than the human, as it may consider the majority potential computable alternatives in a more efficient manner. It comes, therefore, as no surprise, that 63% of the respondents confirm that they would buy a car that takes decisions in such critical situations. The overall positive view of the self-driving car technology per se is in line with existing research and expectations (Rödel et al, 2014; Kyriakidis et al, 2015; Bonnefon et al, 2016).
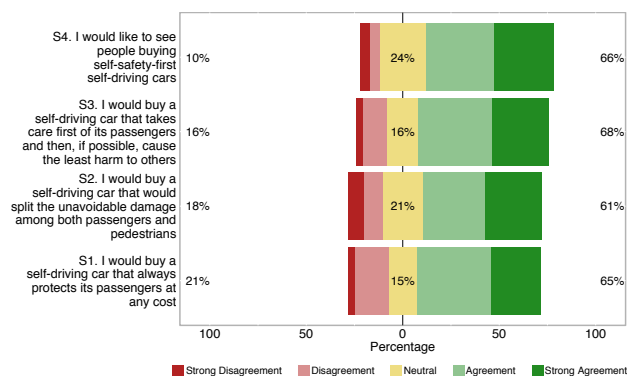

Fig. 3 Survey: Self-Safety

It was hypothesized (H2) that Self-safety prioritization integrated into the decision-making process of the self-driving car may impact its acceptance. From the literature, it is already known that ethical behavior may impact how self-driving cars are perceived. Especially the self-safety mentality, which puts first the safety of the passengers above all others, has also been observed in other studies (Bonnefon et al, 2016). Looking at the details of the respondent answers (shown in Figure 3) it is observed that the overwhelming majority of people would be interested in buying such cars, i.e., cars that take care of the passengers first and then consider alternative options of others (e.g., pedestrians). This seems to be putting trust in technology to protect them and enhance their safety and overall experience on the road. Interestingly a similarly high number of people indicate that they would be interested also in cars that split the unavoidable damage among passengers and pedestrians. In such scenarios, if both passengers and pedestrians, obtain no life-threatening injuries rather than one person being fatally injured and the rest being "saved without a scratch". The latter indi-

cates that people would be willing to accept damages in an effort to minimize overall harm caused during an accident, and are not only focused on egoistic behaviors. In the acquired responses, it is observed that they also would buy only cars that always protect the passengers at any cost. Such observed behavior is contradictory and is a potential indicator that people are open for all options or that they cannot (or do not want to) fully assess the context of damages during unavoidable accidents in discussion. The survey stays at a high level, and here there is potentially room for further investigation, for instance, by using specific scenarios and quantifiable damage (e.g., injury, death, etc.) to get more accurate and consistent behavior. It is reported (Bonnefon et al, 2016) that in such concrete scenarios, e.g., specifying the number of lives saved or lost could pose a differentiator. While this factor captures at a high level the potential issues, a deeper investigation is needed on its parts, e.g., self-preservation, kin-preservation, passenger preservation (Rhim et al, 2020).
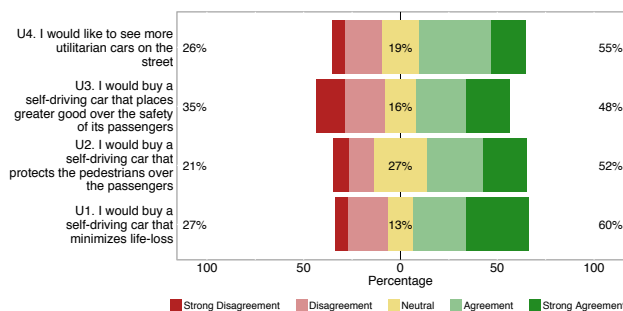

Fig. 4 Survey: Utilitarianism

It was hypothesized (H3) that utilitarianism integrated into the decision-making process of the self-driving car may impact its acceptance. It is worth noting that both the path coefficient and the CR values are very similar for Utilitarianism and Self-Safety, which reflect upon different angles on the ethical aspects in self-driving cars and which are context-wise (and as shown impact-wise) distinct from the technology factor. Utilitarian ethics are in discussion on the general artificial intelligence context, and especially in the context of the self-driving cars. Looking at the details of the respondents' answers (shown in Figure 4), it is observed that generally, people would like to see more utilitarian cars on the street, and generally, they are positive towards buying cars that protect the pedestrians and minimize life-loss overall. However, they seem to be less prone to buying cars that focus only on the greater good over the safety of their passengers. Maybe if all cars impose such a decision (e.g., by regulation), the answers to this

question (U3) would be affected. Generally, the behaviors are in line with what is found in other surveys, as, e.g. by Bonnefon et al (2016), it was also detected that although the people would like to see more utilitarian cars overall on the streets, they probably would not buy themselves one if they could choose. The latter calls for a proper regulatory framework that if choices are available, these should not in any way, lead to direct or implied discrimination of the citizens.
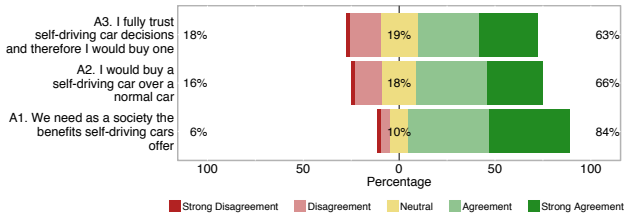

Fig. 5 Survey: Acceptance

Having a look at the questions capturing self-driving car acceptance (shown in Figure 5), it can also be noted that the overall trends detected so far are compatible with what has been found in other surveys. There is a consensus (of 84%) that we do need as a society the benefits that self-driving cars offer. Such benefits seem to be well understood and in line with what people expect, e.g., traffic efficiency, pollution reduction, and elimination of human errors, prevention of accidents, etc. The majority of them (66%) also positively state that they would buy a self-driving car over a normal one. To a lesser degree, though, such buy decisions are made because there is full trust in the decisions that the car might take. It can, therefore, be considered that people might be interested in the perks the self-driving cars might offer (e.g., being able to multitask or not having to park), but when it comes down to specific behaviors that span the spectrum of self-safety to utilitarianism, then things are not (yet) fully clear.

Of interest would be to see if the car passenger behavior changes and if they still consider themselves responsible (indirectly) if the car takes all decisions. Or maybe then an accident is always thought of as "somebody else's fault", which might limit people's empathy and risk identification. On the pedestrian side, would they then also cross the streets more carelessly, as they know that the self-driving cars deal better than humans? Would trust in machines taking the best possible decisions increase? As such, after some years of symbiosis with self-driving cars, would anyone question if better decisions could have been taken by the car? Approaches that attempt to simulate actions and predict their consequences have been proposed (Vanderelst and Winfield, 2018), but critical situations put an extra requirement of real-time decision making, which pushes further the challenge. Due to the increased complexity and emergence of artificial intelligence in the self-driving cars, would it still be possible to link a specific decision taken during an accident and the conditions that derived it? If not, how could then mistakes be pinpointed and corrected?

For users to adopt self-driving cars, they need to trust them. This implies trust in their decision processes, as well as trust in the timely execution of decisions taken. This calls for more transparency in the modern machine learning technologies and their utilization that empowers contemporary efforts of self-driving cars. This is a challenging issue, since for the car to make the right decision, it will have to trust its contextual awareness (Fernandez-Rojas et al, 2019) and its sub-components.

In addition, designers, engineers, and technologists, need to understand better and rethink the paradigm of traffic management and accident management, including avoidance. As an example, it would be very limiting to further consider the envisioned hyper-connected self-driving cars as in the operational context of legacy cars, operating as singletons and relying only on their sensors and internal logic, without taking advantage of their hyper-connectivity and higher-level skills such as coordination and negotiation. The self-driving car of the future needs to act as part of a system of systems and operate within a context that is defined by the constant interaction with other stakeholders, e.g., other self-driving cars, intelligent infrastructure, localized services, traffic management systems, etc. For instance, in an unavoidable accident involving two self-driving cars, the cars may attempt to negotiate and sync their actions so that they collectively minimize damages. In urban environments, self-driving cars can consider pedestrian and environmental factors (Rasouli and Tsotsos, 2019), analyze in real-time their behavior, and may even attempt to influence it. Understandably this increases automotive software complexity (Vdovic et al, 2019), but it will enhance the car's capabilities, and this collective intelligence may benefit the individual actors (self-driving cars) to make better and more timely decisions, and therefore the public overall.

While the participant sample ($n = 62$) is sufficient and other empirical studies also have similar samples, e.g., $n = 70$ (Rhim et al, 2020), the sample is not probabilistic and as such it is also suggested to have a larger and more diverse sample that covers all additional aspects to be investigated. Nevertheless, for the carried out investigations, as the statistical analysis has shown,

this sample is sufficient to make the correlations discussed, and pose a starting point for future research.

While ethics play a role, there are several ethical frameworks that could be deployed to self-driving cars. For instance, Relativism, Utilitarianism, Absolutism, Deontology, Pluralism, just to name a few, also have an impact on the self-driving car acceptance (Karnouskos, 2020). However, it is not clear if all the cars should have the same ethics or different ones, as well as if the driver should be able to prescribe them. Could such decisions be then influenced by the laws of a specific country, and would then illegal markets arise to modify the expected car behavior with another one (due to a user request or as a result of hacking)? In such cases, liability issues become even more complicated. Furthermore, since real-time decisions need to be made, and this may depend on hardware and software speed, what would be the min requirements for self-driving car decision-making? Would it also then be ethical to sell cars with cheaper components that may result in being too slow to reach optimal decisions in critical situations effectively? The latter implies that technology interferes with ethics, and the cost is injected as a factor for class discrimination since the rich may be able to afford electronics that take faster/better decisions than the more impoverished citizens.

While we have investigated the acceptance of self-driving cars from technology and ethical perspective, there are also other factors, e.g., law, regulation, culture, etc. (Li et al, 2019; Shladover and Nowakowski, 2019; Rhim et al, 2020) that need to be considered. Such non-technical aspects affect moral reasoning as for instance, cross-cultural comparison between Korea and Canada has recently shown (Rhim et al, 2020). Such factors need to be investigated, and their interplay with ethics and technology needs to be assessed both as standalone (in-depth) as well as part of an ecosystem (horizontally).

This work is based on a limited set of empirical data gathered via a survey. While the sample is sufficiently analyzed with the methods used (SEM), bias may exist in the answers received in the survey. In addition, several other factors such as technical expertise, cultural aspects, social expectations, etc. may have influenced the answers of the respondents. As such, this work should be seen more as an effort to show some indications and discuss pertinent issues relating to utilitarianism, self-safety, and technology, but should not be generalized as more focused and larger-scale investigations need to be made.

Finally, despite the vivid ongoing discussions in literature, there is also the viewpoint that the car should never reach a point of making a moral decision. While "trolley" dilemmas can be useful in philosophy and psychology, in the real world, they are hard to detect and hard to act, and therefore utilizing "trolley" dilemmas to train self-driving cars on how to act, may simply be an engineering and policy distraction (Freitas et al, 2019). If such a situation arises, then maybe the self-driving car should do its best to maintain its predictability, e.g., trajectory, and rely on other external stakeholders for taking life/death decisions. This, in practice, would mean that a car in a collision course with pedestrians should brake and keep a straight trajectory so that the pedestrians can anticipate the car's behavior and act themselves to get out of harm's way.

5 Conclusions

Mass production and operation of self-driving cars for personal or commercial purposes are underway. However, the delegation of responsibilities to the self-driving car is vague when it comes to the area of unavoidable accidents. In that case, an algorithmic decision needs to be made that will eventually harm the passengers, others (e.g., pedestrians), or both. The implications on the acceptance of self-driving cars that will exhibit behaviors such as protecting the common good by causing the least harm (utilitarianism) or protecting its passengers first and then the rest (self-safety first), although discussed are still not clear. Ongoing research has identified several factors and some of their implications. Three key such factors, i.e., Technology, Self-Safety, and Utilitarianism, were hypothesized to be linked to the self-driving car acceptance. The evaluation of the survey data shows that all three factors contribute (in a statistically significant way) to self-driving car acceptance, with technology being the major contributor, followed by the other two factors, i.e., utilitarianism and self-safety which seem to have comparable contributions. It is of high importance to understand how self-driving car acceptance can be influenced, since as also seen in the survey, and there is an overwhelming interest in investing in the benefits they bring. However, some behaviors lead to paradoxes, e.g., people want others to have utilitarian cars, but they would prefer to potentially buy self-safety cars (Bonnefon et al, 2016), which will end up with more self-safety cars on the streets and that would contradict the global preference for utilitarian cars. While this work has brought to attention some elements of the interplay of self-driving car behavior and their acceptance, there is evident the need for more in-depth analysis along also other factors such as law, regulations, culture, power, societal injustice, etc., just to name a few.

References

Bansal P, Kockelman KM (2016) Are we ready to embrace connected and self-driving vehicles? a case study of texans. Transportation DOI 10.1007/s11116-016-9745-z

Bansal P, Kockelman KM, Singh A (2016) Assessing public opinions of and interest in new vehicle technologies: An austin perspective. Transportation Research Part C: Emerging Technologies 67:1–14, DOI 10.1016/j.trc.2016.01.019

BBC (2016) Tesla car drives owner to hospital after he suffers pulmonary embolism. URL https://goo.gl/CimvFn

Bonnefon JF, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. Science 352(6293):1573–1576, DOI 10.1126/science.aaf2654

Borenstein J, Herkert J, Miller K (2017) Self-driving cars: Ethical responsibilities of design engineers. IEEE Technology and Society Magazine 36(2):67–75, DOI 10.1109/mts.2017.2696600

Bremner P, Dennis LA, Fisher M, Winfield AF (2019) On proactive, transparent, and verifiable ethical reasoning for robots. Proceedings of the IEEE DOI 10.1109/jproc.2019.2898267

Carsten P, Andel TR, Yampolskiy M, McDonald JT (2015) In-vehicle networks. In: Proceedings of the 10th Annual Cyber and Information Security Research Conference on - CISR '15, ACM Press, DOI 10.1145/2746266.2746267

Coca-Vila I (2017) Self-driving cars in dilemmatic situations: An approach based on the theory of justification in criminal law. Criminal Law and Philosophy DOI 10.1007/s11572-017-9411-3

Ethik-Kommission (2017) Automatisiertes und vernetztes fahren. Tech. rep., Federal Ministry of Transport and DIgital Infrastructure, Germany, URL http://www.bmvi.de/bericht-ethikkommission

Fernandez-Rojas R, Perry A, Singh H, Campbell B, Elsayed S, Hunjet R, Abbass HA (2019) Contextual awareness in human-advanced-vehicle systems: A survey. IEEE Access 7:33304–33328, DOI 10.1109/access.2019.2902812

Freitas JD, Anthony SE, Alvarez G (2019) Doubting driverless dilemmas. PsyArXiv Preprints DOI 10.31234/osf.io/a36e5

Guo J, Kurup U, Shah M (2019) Is it safe to drive? an overview of factors, metrics, and datasets for driveability assessment in autonomous driving. IEEE Transactions on Intelligent Transportation Systems pp 1–17, DOI 10.1109/tits.2019.2926042

Haboucha CJ, Ishaq R, Shiftan Y (2017) User preferences regarding autonomous vehicles. Transportation Research Part C: Emerging Technologies 78:37–49, DOI 10.1016/j.trc.2017.01.010

Hevelke A, Nida-Rümelin J (2014) Responsibility for crashes of autonomous vehicles: An ethical analysis. Science and Engineering Ethics 21(3):619–630, DOI 10.1007/s11948-014-9565-5

Hicks D, Simmons R (2019) The national robotics initiative: A five-year retrospective. IEEE Robotics & Automation Magazine DOI 10.1109/mra.2019.2912860

Hohenberger C, Spörrle M, Welpe IM (2016) How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups, vol 94. Elsevier BV, DOI 10.1016/j.tra.2016.09.022

Hussain R, Zeadally S (2019) Autonomous cars: Research results, issues, and future challenges. IEEE Communications Surveys & Tutorials 21(2):1275–1313, DOI 10.1109/comst.2018.2869360

IEEE (2018) Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. URL http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Johannesson P, Perjons E (2012) A Design Science Primer. CreateSpace

Karnouskos S (2020) Self-driving car acceptance and the role of ethics. IEEE Transactions on Engineering Management 67(2):252–265, DOI 10.1109/tem.2018.2877307

Karnouskos S, Kerschbaum F (2018) Privacy and integrity considerations in hyperconnected autonomous vehicles. Proceedings of the IEEE 106(1):160–170, DOI 10.1109/jproc.2017.2725339

Kyriakidis M, Happee R, de Winter J (2015) Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation Research Part F: Traffic Psychology and Behaviour 32:127–140, DOI 10.1016/j.trf.2015.04.014

Li S, Sui PC, Xiao J, Chahine R (2019) Policy formulation for highly automated vehicles: Emerging importance, research frontiers and insights. Transportation Research Part A: Policy and Practice 124:573–586, DOI 10.1016/j.tra.2018.05.010

Lin P (2015) Why ethics matters for autonomous cars. In: Autonomes Fahren, Springer Nature, pp 69–85, DOI 10.1007/978-3-662-45854-9_4

Malle BF, Scheutz M, Arnold T, Voiklis J, Cusimano C (2015) Sacrifice one for the good of many?: People apply different moral norms to human and robot agents. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15, ACM, DOI 10.1145/2696454.2696458

Mordue G, Yeung A, Wu F (2020) The looming challenges of regulating high level autonomous vehicles. Transportation Research Part A: Policy and Practice 132:174–187, DOI 10.1016/j.tra.2019.11.007

Nees MA (2016) Acceptance of self-driving cars: An examination of idealized versus realistic portrayals with a self- driving car acceptance scale. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 60(1):1449–1453, DOI 10.1177/1541931213601332

NTSB (2019a) Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018. URL https://ntsb.gov/investigations/AccidentReports/Reports/HWY19FH008-preliminary.pdf

NTSB (2019b) Highway Preliminary Report: HWY19FH008. URL https://ntsb.gov/investigations/AccidentReports/Pages/HWY19FH008-preliminary-report.aspx

PWC (2016) Driving the future: understanding the new automotive consumer. Tech. rep., Price Waterhouse Coopers (PWC), URL https://tinyurl.com/ugx334m

Rasouli A, Tsotsos JK (2019) Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems pp 1–19, DOI 10.1109/tits.2019.2901817

Rhim J, Lee Gb, Lee JH (2020) Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. Computers in Human Behavior 102:39–56, DOI 10.1016/j.chb.2019.08.010

Rödel C, Stadler S, Meschtscherjakov A, Tscheligi M (2014) Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. In: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications – AutomotiveUI '14, ACM, DOI 10.1145/2667317.2667330

Shladover SE, Nowakowski C (2019) Regulatory challenges for road vehicle automation: Lessons from the california experience. Transportation Research Part A: Policy and Practice 122:125–133, DOI 10.1016/j.tra.2017.10.006

de Sio FS (2017) Killing by autonomous vehicles and the legal doctrine of necessity. Ethical Theory and Moral Practice DOI 10.1007/s10677-017-9780-7

Sütfeld LR, Gast R, König P, Pipa G (2017) Using virtual reality to assess ethical decisions in road traffic scenarios: Applicability of value-of-life-based models and influences of time pressure. Frontiers in Behavioral Neuroscience 11, DOI 10.3389/fnbeh.2017.00122

Vanderelst D, Winfield A (2018) An architecture for ethical robots inspired by the simulation theory of cognition. Cognitive Systems Research 48:56–66, DOI 10.1016/j.cogsys.2017.04.002

Vdovic H, Babic J, Podobnik V (2019) Automotive software in connected and autonomous electric vehicles: A review. IEEE Access 7:166365–166379, DOI 10.1109/access.2019.2953568

Zmud J, Sener IN, Wagner J (2016) Consumer acceptance and travel behavior impacts of automated vehicles. Tech. rep., Texas A&M Transportation Institute, URL http://tti.tamu.edu/documents/PRC-15-49-F.pdf, pRC 15-49 F