Self-driving Car Acceptance and the Role of Ethics

Stamatis Karnouskos

***Abstract*—Mass availability of self-driving cars is ante portas and independent of their sophistication, unavoidable fatal accidents will occur, where the car will have to take life and death decisions. However, there is a knowledge gap since the impact, the ethical frameworks utilized in the decision-making process, have on the overall acceptance of self-driving cars, is not well-investigated. This work is addresses the key question: In the scope of unavoidable accidents, what is the effect of different ethical frameworks governing self-driving car decision-making, on their acceptance? This quantitative positivist research investigates the link of selected ethical frameworks, i.e., Utilitarianism, Deontology, Relativism, Absolutism (monism), and Pluralism, to the acceptance of self-driving cars. It is hypothesized that they have an impact on the acceptance of the self-driving cars, and a model linking them to it is proposed and assessed. All five selected ethical frameworks investigated are found to have an effect on self-driving car acceptance, which implies actions for several involved stakeholders as these may be a deciding factor for the success or failure of the self-driving car market introduction.**

***Index Terms*—Autonomous vehicles, machine ethics, ethical dilemmas, self-driving car acceptance.**

I. Introduction

Artificial Intelligence (AI) has made leap steps in the last couple of years towards its practical applications in different domains, including that of intelligent transportation, where the introduction of self-driving cars is a game-changer [1]. Today the majority of car manufacturers work on autonomous driving while some have commercially released self-driving cars (with various degrees of sophistication), while more than 10 million of them are expected to be available by 2020 [2]. The promise of self-driving cars comes with significant benefits for the individuals and society [3]–[5]. However, delegating driver's responsibilities to a "robot on wheels", has far-reaching implications, especially when it comes down to automated decision-making in critical situations.

Human errors are the main cause of vehicle crashes, e.g., 94% in U.S. [6]. Self-driving cars bear the promise to significantly reduce accidents by taking the human factor out of the equation [3], while in parallel monitor the surroundings, detect and react immediately to potentially dangerous situations and driving behaviors. However, that is not a guarantee as the recent Uber test vehicle accident attests [7], where a combination of factors led to a pedestrian being killed by a self-driving car. Independent of technological progress, some accidents will still be unavoidable [8] and involve endangerment of human lives, either in the self-driving car (e.g., driver, passengers) or outside of the car, e.g., pedestrians, and other cars. As unavoidable accidents are defined those critical situations, where for the reaction time available (to a human or self-driving car), no solutions can be found, that may avoid an accident. In this work, the focus is on unavoidable accidents that lead to fatalities, i.e., accidents where human lives are lost. In the unavoidable accident context, a self-driving car must decide what action to take; and this implies also deciding on the potential casualties [8,9], something that is only partly investigated today legally and ethically.

The ethical dimension of self-driving car decision-making in critical situations is evident. Overall ethical dilemmas in critical situations are not a new domain, and probably the most well-known is the "trolley problem" [10]. Apart from the general question if and how ethics can be programmed [11] or standardized [12], in self-driving cars what is of high interest and is not sufficiently investigated, is the impact of ethics on the acceptance of self-driving cars. Hence, the research question investigated [13] can be posed as: *In the scope of unavoidable accidents, what is the effect of different ethical frameworks governing self-driving car decision-making, on their acceptance?*

To exemplify the ethics impact [14,15] on the acceptance of self-driving cars, one has to consider the situation of an eminent fatal accident involving pedestrians and car passengers. En route, the self-driving car identifies that it has a sudden brake failure and needs to decide whether to protect its passengers from harm, although that would result in the death of some pedestrians or crash into a wall, which would protect pedestrians, albeit kill some of its passengers. The decision taken by the self-driving car has an ethical dimension, and the question that arises is what sort of

S. Karnouskos is with SAP, Dietmar Hopp Allee 16, D-69190, Walldorf, Germany, (email: karnouskos@ieee.org). Work carried at Stockholm University, Department of Computer and Systems Sciences, SE-16407 Kista, Sweden.

ethics guide the car's decision algorithms, and what are the overall implications. For instance, one could argue that innocent passengers ought to be spared, and hence the car passengers should bear the risk of being fatally injured. This most probably would be seen positively by the majority of the people in a city, especially the non-drivers. However, the question that is raised is if anyone would then buy such a car if s/he knows s/he is in high danger; probably not. Subsequently, that may result in a decrease in the sales of self-driving cars, and they will never reach a critical mass. Hence, the envisioned benefits coupled with their existence (e.g., overall reduction of accidents) would also not be materialized as expected.

As it can be seen, the ethics embedded in the decision-making of a self-driving car, especially in the case of unavoidable accidents, would most probably impact their acceptance by the public. Initial research indicates that this is indeed the case [16]–[20]. Also, the nature of the ethics, i.e., the ethical framework utilized may also play a role, something that is not sufficiently investigated. The problem pertaining this work is to identify if the existence of a specific ethical framework in the decision-making process of a self-driving car, especially in critical situations where life and death decisions are made, has an impact to its acceptance [16]–[19]. As there are several ethical frameworks available [21], that signify specific ways moral decisions are made, the question that arises is which of these may impact the self-driving car acceptance. Therefore, ethics pertaining to decision-making processes in self-driving cars is a critical research issue with practical implications and needs to be addressed. As it is pointed out [17]: "Figuring out how to build ethical autonomous machines is one of the thorniest challenges in artificial intelligence today. As we are about to endow millions of vehicles with autonomy, a serious consideration of algorithmic morality has never been more urgent."

Acceptance of self-driving cars is a major issue that can decide over success or failure of their introduction [1,16,22,23]. A lot of research is devoted to the qualitative aspects pertaining to the benefits or challenges of self-driving cars and their potential impact on acceptance, but without an equal focus on quantitative empirical research. Although some empirical surveys exist and identify acceptance as a key issue for the success of self-driving cars [24]–[45], the ethical aspects are hardly considered, and when they do, they are indirect or mostly related to privacy. Recently, some research focused explicitly on the ethics of self-driving cars and the impact on their acceptance [17]–[19]. Such focused surveys constitute a significant step towards addressing the ethical aspects pertaining self-driving cars. However, most of them discuss some example scenarios pertaining to one ethical framework (usually utilitarian) and not considering sufficiently other frameworks.

The pertinence of the issue is also highlighted in governmental reports; for instance, in Germany proposals are made [5] towards issues such as property damage vs. personal injury, utilitarian decisions (based on personal features such as age, gender, physical or mental constitution), liability, data management, etc. As it can be seen, there is a need to investigate the potential effect of ethical frameworks on the acceptance of self-driving cars. The main contribution of this work is the investigation of the hypothesized relationship of diverse ethical frameworks and self-driving car acceptance, as this has the potential to generate new insights [13]. A model assessing the effect of ethical frameworks in self-driving car acceptance is hypothesized and assessed via the collected empirical data.

The rest of the paper is structured as follows: the overall approach is presented in section II, while the main ethical frameworks and their link to the investigated area is discussed in section III Subsequently, the empirical results collected are presented in section V, while their implications are discussed in section VI. Finally, several additional challenges and research directions are presented in section VII.

II. Approach

There are several methods that can be used to approach the research question posed, both of qualitative [46] and quantitative form [47]. The qualitative form may be fit for identifying the factors, and potentially come up with new ones that are not in the literature. However, since the focus is set explicitly to the selected factors that are relevant to the digital era, the identification of new factors (ethical frameworks) is not in the context of this work. Besides, since the aim is to measure the impact of the selected factors identified, i.e., Utilitarianism, Deontology, Relativism, Absolutism (monism), and Pluralism to the acceptance of self-driving cars, a common practice is to do this via statistics on quantifiable data. Therefore, the quantifiable research is seen as more appropriate to follow in order to answer the posed research question.

In this work quantitative positivist research [47] is carried out, and the empirical data is collected via a questionnaire. The larger ethics area is covered by experimental philosophy, and similar approaches as the one used here for capturing and quantification exist, e.g., the Ethics Position Questionnaire (EPQ) [48]. This choice is also justified by what other researchers have used in similar contemporary studies [17,25]. In addition,

Preprint version of doi:10.1109/TEM.2018.2877307

some quantitative surveys dealing with self-driving car acceptance are available [17]–[19], which however do not reflect upon all ethical frameworks considered in this work. As such, it was seen as necessary, to construct a new survey, which addresses the needs of this work, while in parallel considering the previous quantitative research.

With respect to the process followed, first, the ethical frameworks are selected and described. Ethical frameworks are posed in the unavoidable accident context and a model that hypothesizes their link to the acceptance of self-driving cars is proposed. Subsequently, a survey with questions that capture the identified factors (ethical frameworks) is constructed and empirical data is collected. The sampling frame is general, the initial scope is university students (at Master's level) as they pose a good mix of technology savviness and will be able to easily understand the context in which self-driving cars will have to operate. This technique is seen as effective [49], especially in small-scale research projects.

For the analysis, common statistics processes are used e.g., descriptive statistics, factor analysis etc. [49]. Descriptive statistics of the data collected enable understanding of the nature of the dataset as well as its quality. More advanced tests are done, i.e., Kaiser-Meyer-Olkin Measure of Sampling Adequacy, Bartlett's Test of Sphericity, Cronbach's α, χ^2 etc. to justify aspects such as adequacy and internal validity. Factor analysis is used as it examines the inter-correlations that are evident between the items (the survey questions) and reduces them to smaller groups known as factors (ethical frameworks). Subsequently, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), i.e. SEM [50] are performed in order to investigate the link between the hypotheses and the impact on self-driving car acceptance. SEM is a special form of CFA, that enables specification of factors and variables of causal in nature. The results are then critically discussed and analyzed.

There are several ways to demonstrate instrumentation validity [51]. Through careful selection of factors and unambiguous questions that capture all aspects of the factors, we aimed at construct validity. Since SEM is utilized, construct validity is also addressed as "in short, SEM, including both least-square and covariance-based techniques, accounts for error, including measurement error, in a holistic way" [47]. For reliability, also Cronbach's α as an indicator of internal validity [52] is measured. In addition, reliability is enhanced if respondents give comments [53]; hence this kind of feedback is integrated into the survey. The data is collected anonymously, and hence we do not foresee any "effect of the researcher".

As the area of intelligent machine-automated decision-making is vast and complex, several delimitations are made. This study limits itself to a subset of the ethical frameworks pertinent to the digital era [21], i.e., utilitarianism, deontology, relativism, absolutism (monism), and pluralism. In addition, although the potential impact of ethical frameworks in self-driving car acceptance at large is investigated, several fine-grained issues and satellite questions, are not in the eminent scope of this work, e.g., the relationship between the self-driving car's ethics and the driver's is of interest, especially if these two do not match. Even for the selected frameworks, this study stays knowingly at a relatively high level, in order to make a first assessment of their discrete effect on self-driving car acceptance. Hence it does not consider the variations within a specific framework, e.g., within the Utilitarian framework, fine-grained aspects such as the viewpoints of western or Chinese philosophy, negative utilitarianism, motive utilitarianism, preference utilitarianism, etc. are not considered. Furthermore, this work does not investigate fine-grained aspects of the respondents and their impact, e.g., cultural, geographical, financial, religious, gender, societal, etc. aspects, which are seen as future research.

III. Ethical Frameworks

Ethics is one of the branches of philosophy, and there are multiple fine-grained ethical frameworks to be considered. However, in this work, it was decided to follow the classification that focuses on the digital era at large [21], which includes self-driving cars. The following frameworks were selected as representative: Utilitarianism, Deontology, Relativism, Absolutism (monism), and Pluralism.

A. *Utilitarianism*

Utilitarianism is a normative ethical framework that considers as the best action, the one that maximizes a utility function by considering the positive and negative consequences of the choices pertaining to the decision. It is a form of consequentialism, where the decision maker tries to think of all possible good or bad consequences of acts, and then by weighting them against each other, to determine which action will generate the most positive outcome [21]. Therefore, the introduction of a function that calculates the pleasure vs. pain (good vs. bad), referred to as hedonic calculus, can guide decision-making. Assuming for instance that one can somehow model the consequences of actions and uncertainties, rational decision-making methods can be applied [54,55],

Preprint version of doi:10.1109/TEM.2018.2877307

alternatives can be calculated, and the best one can be selected.

Utilitarianism is a cluster of related theories developed over several decades, and therefore there are several criticisms against it. A key question is how to weight the possible outcomes. Using a purely quantitative approach would result in summing positive and negative utils, but not everything can be measured in terms of costs (pain) or benefits (pleasure) and therefore the quantification in a hedonic calculus, may be impossible to lead to the establishment of a standardized scale. Apart from the difficulty in modeling, quantification, and assessment (number of utils), other aspects are also challenging. For instance, how far in the future should such consequences be considered? As uncertainty grows in the future, this impacts the utility function and therefore has an impact on the reliability of the decision. Furthermore, for whom should such consequences be assessed? For the individual, his/her family, the society?

In Utilitarianism, the end justifies the means. In history, there are many examples where decisions were taken and justified under a utilitarian framework. For instance, during a war, the sacrifice of some people (soldiers) is justified to protect the rest. Similarly, loss of freedom and rights for some (slaves) is justified in utilitarianism if there is a benefit (positive utility function) for the many (society). In modern times, compared to humans, robots are expected to take utilitarian actions, that sacrifice one person for the good of many, and are more blamed when they do not [56].

B. Deontology

Deontology is derived from the Greek "deon" (=duty) and "logos" (=science). Deontology is a normative ethical framework and considers that there are rules that have an absolute quality in them, which means that they cannot be overridden. As such, deontologists reject that what matters are the consequences of an action, and focus that what matters is the kind of action to be taken. Therefore, a deontologist would always have to "do the right thing for the right reason, because it is the right thing to do", i.e., what matters is doing his/her duty.

Typical examples are evident in religions, e.g., in religious pacifists where life is sacred, violence against others is rejected, by being wrong. Hence a religious pacifist would reject killing another human, no matter the consequences, even if this results in losing one's own life [21]. Such behaviors are justified by the respondents of a survey [19] in that "they never could live with the knowledge of being (somehow) responsible for the death of others".

Rational deontology is reflected by Immanuel Kant's categorical imperative "Act only on that maxim through which you can at the same time will that it should become a universal law" [57]. This determines what the moral duties are and poses an absolute and unconditional rule that must always be obeyed independent of the consequences. As such, Kant's categorical imperative can be seen as "a kind of procedural way of determining what actions are right" [21]. In deontology, the emphasis is put on the act itself – in contrast to utilitarianism where the emphasis is on the outcome of the action and how right it is. In deontology, other persons are treated as an end in itself, but never as means. Hence it is considered that there are ethical absolutes, such as human rights, which are treated as universal values. In literature, there are arguments that there is a need to "program the crash algorithms of autonomous cars based on a deontological understanding of the system of justifications in criminal law" [58].

Due to its absolute nature, deontology is criticized as it always prioritizes right over good. For instance, lying is always wrong according to Kant, even if something good can be derived. However, such a deontological position is not always justified, and people cannot always accept it, especially if it involves saving lives.

C. Relativism

Ethical Relativism is a meta-ethical framework where it is argued that "all norms, values, and approaches are valid only relative to (i.e., within the domain of) a given culture or group of people" [21]. Ethical relativism emerges as there is a range of practices that although considered morally acceptable in some societies, they are condemned by others. Hence, in this framework, it is proposed that a society's practices can be judged only by its own moral standards. This framework offers two advantages: (i) tolerating the views and practices of "others", since such exist and (ii) it offers a relief (or better said an excuse) in the sense that if everything is relevant to a specific culture or group, then one does not need to exhaustively search for globally valid values, practices or frameworks. However, following ethical relativism, tolerance appears "to emerge as itself a universally valid ethical norm or value" which contradicts the fundamental ethical relativism, that states that since all values are relative, tolerance should also be relative and not universally valid [21]. In addition, a logical fallacy of affirming the consequent is in place, i.e., if there are no universal values, then diverse values are to be found; diverse values are found; therefore, no universal values exist.

Preprint version of doi:10.1109/TEM.2018.2877307

Ethical relativism argues that ethics are a product bound to a culture, time etc. which may lead to acceptance of values or practices because that is the way it was done at the time, or within a community. This impacts moral beliefs, as an action is right or wrong according to that society's norms, and therefore one has to obey these norms, as any deviation would be considered immoral. Ethical relativism would lead to moral paralysis as no ethical judgment can be done against the "others". For instance, the apartheid in South Africa or the treatment of the Jews in Nazi Germany cannot be condemned, as these were the values of that society; which is unacceptable. Furthermore, it promotes isolation among cultures, as it is argued that ethical values from the "others" cannot be learned. As a gist, ethical relativism brings up the issue of difference between moral aspects that depend on culture, time, etc.

D. Absolutism (monism)

Ethical absolutism or ethical monism is a meta-ethical framework that is on the antipodal point of the ethical relativism. This framework, also referred to as "doctrine of unity", can be described as follows: "There are universally valid moral rules, norms, beliefs, practices, etc. [...that] define what is right and good for all at all times and in all places – those that differ are wrong" [21]. This position implies that an ethical absolutist knows clearly those universally valid moral rules, norms, beliefs, practices, and any deviation or difference is therefore wrong or invalid.

As such, for instance, beliefs that agree with the views of an absolutist are praised, while those that differ are condemned. As absolute positions can be taken for or against a matter, it may well be that two absolutists have different views on the same aspect, each considering his/her view as right and the other as wrong. In comparison, an ethical relativist would not condemn any of the two (as none of them can claim universal validity), but consider tolerance of the ethical differences, since it's all a matter of culture, personal preferences etc.

E. Pluralism

Ethical pluralism is a meta-ethical framework that rejects absolutism (that there is only one correct moral truth) and relativism (that there is no correct moral truth) as unsatisfactory and proposes that there is a plurality of moral truths. It is sometimes referred to as "doctrine of multiplicity". The ethical pluralist argues that indeed there are universal values (as indicated in absolutism) however, instead of considering that there is only a single set always applicable, it considers that there are many which can be interpreted, understood and applied in diverse contexts (as indicated in ethical relativism). As such, in ethical pluralism differences can be tolerated (as a relativist would do), rather than condemned (as an absolutist would do), and practices that violate a basic value (e.g., human rights) can be condemned as immoral (which was not possible with the ethical relativism). Hence "ethical pluralism allows us to see how people in diverse cultures may share important norms and values; but at the same time, we are able to interpret and apply these norms and values in sometime very different sorts of practices – ones that reflect our own cultural contexts and traditions" [21]. Ethical pluralism goes beyond the western tradition and appears to be a widely spread approach in approaching ethical differences, while it can also be utilized for globally recognized issues such as privacy and how it is protected by laws all over the world.

IV. Linking Ethics & Self-driving Car Acceptance

The key question is if the ethical frameworks have an impact on the acceptance of self-driving cars, especially in the context of unavoidable accidents where life and death decisions need to be made by it. There are several satellite issues such as if people trust the car to take the decision, if the car can actually take a better decision than the human driver, or if the driver ought to be given back the control in critical situations. Subsequently, a key unresolved issue is that if the self-driving car takes the decision, which entity is then liable, etc. Such questions are still not sufficiently addressed, and the ethical aspects are complex especially considering that the self-driving car may act as a moral proxy [59] of its users. As with all human-related research that carries ethical and social dilemmas, the concern is also how much responsibility delegation can be realized or shared among humans and machines [60,61].

There are several factors that play a key role on how people think about self-driving cars [8,23]–[45,62,63]. The question that arises is on the ethical expectations, and the empirical quantitative research tackling this aspect is limited [17]–[20]. For instance, recent research [56] indicates that utilitarianism is expected from machines, and that "simple value-of-life models approximate human moral decisions well" [20].

Should all the self-driving cars have utilitarian ethics, and in the case of an unavoidable accident, follow a least-harm strategy? But then for whom? For the car's passengers, the others (e.g., pedestrians, other car's passengers), or for all? Would actually then people buy such cars or would they prefer to stick to the old technology (where the driver is in full-control)? Or is there an Aristotelian

Preprint version of doi:10.1109/TEM.2018.2877307

golden mean, e.g., where the humans have to take that decision, and if so, can they actually be better or worse at it? Or maybe deontological ethics ought to be in place so that absolute values are imprinted and the self-driving car has always to follow?

As an example, some scholars [64] propose that a mandatory ethics setting should be enforced on all self-driving cars, rather than individual ethics (even when these have limitations), as the society is on the long run better-off. If not all cars are built with utilitarian ethics, can or should the driver (or other passengers of the car) choose the ethics of the owned "moral proxy" [59], i.e., her/his car and then decide if s/he wants it to take non-utilitarian decisions, e.g., protect the passengers at all cost (self-safety first) or even share the harm caused among the passengers and third-party victims? Would that variety of ethical alternatives enable the self-driving cars to be better accepted in future society, e.g., motivate people to buy them? Would there be any side-effects, e.g., the creation of a black market where the technology-specialists (or those who can afford to pay them) would have their cars rigged to deviate from any standard profiles (e.g., turn a utilitarian car to a self-safety-first car)? As such, investigation of key ethical aspects is relevant within the context of self-driving cars and could provide some indicators.

Machine-related ethical aspects are a key issue [65], and with self-driving cars, there is a need to integrate ethics in the decision-making algorithms of the self-driving car [66]. However, this is challenging and especially in life and death dilemmas, e.g., killing the car passengers vs. killing pedestrians, experimental ethics come into play [56] and impact overall acceptance of self-driving cars. For instance, research [56] shows that people may accept utilitarian self-driving cars, but may not buy them. This though, creates a paradox, as then the number of self-driving cars on the street would be limited and the benefits to society (e.g., accident reduction) would not materialize. Generally, it is pointed out that a more systemic view, that recognizes systems and relationships, is needed to address ethical concerns [67]. Other research [19] points out that knowing the exact risks of the decision taken influences also the processes as people would be more likely to sacrifice others if their own probability of survival is very low. In addition, as self-driving cars will crash, questions of liability as well as what law might permit or not, arise [23,58,62].

The situation is even more complicated in light of technology concerns and indirect ethical aspects that pertain to privacy, security, and trust in self-driving cars [68]. As security and privacy aspects [69] directly affect the operation of the car [9] or its supporting infrastructure [70], malicious (remote) manipulation, and misuse [71] could in some scenarios bypass the decision processes (and ethics) of a car. That would imply, that the final behavior of the self-driving car, might contradict its original ethics, as now these are bypassed and under the influence of a third-party, who can enforce its own rules.

As it can be seen, this area is of increasing interest to multiple stakeholders, and especially car technology providers who need to accommodate ethical aspects in their algorithms, manufacturers who need to understand the impact of different factors on the acceptance of the future self-driving cars, legislators who need to clarify the operational context [72] of such cars, etc. At the bottom line, how ethics and self-driving cars are treated also "depends upon whether society comes to view these machines as simply more capable cars or robots with their own sense of agency and responsibility" [55].

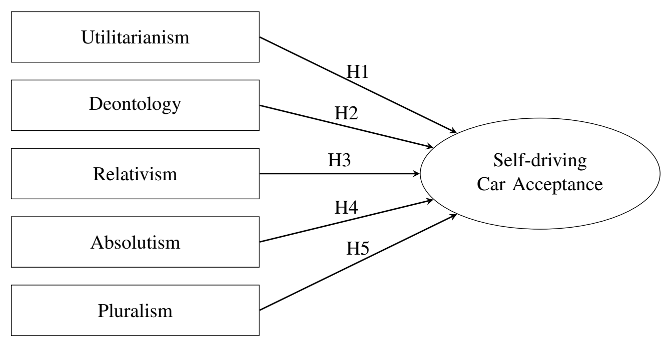

Figure 1. Hypothesized model linking selected factors to the self-driving car acceptance

Five factors (ethical frameworks) are selected from the literature, i.e., Utilitarianism, Deontology, Relativism, Absolutism (monism), and Pluralism. It is hypothesized that the five selected factors (ethical frameworks) may have an effect the acceptance of self-driving cars, as shown in Figure 1, and are analyzed as discussed in section II.

V. Empirical Results & Analysis

A. Data Overview

Empirical data was collected via a survey where each factor was represented with four questions, i.e., Utilitarianism (U1–U4), Deontology (D1–D4), Relativism (R1–R4), Absolutism (A1–A4), Pluralism (P1–P4), and Acceptance of self-driving cars (AC1–AC4). Data were as collected in a 5-level Likert scale [73], ranging from strongly disagree (coded as 1) to strongly agree (coded as 5) for each question in the survey. In total $N = 126$ responses were received, where 37.3% are females and

Preprint version of doi:10.1109/TEM.2018.2877307

62.7% males. From the age perspective, the distribution is 58.7%, 36.5% and 4.8% for 18–29, 30–49, and 50+ years old respectively.

All of the variables are based on the Likert scale, and therefore there is no reason to exclude variables on skewness unless they exhibit no variance. Hence, the focus is on the kurtosis, where a kurtosis > 1 or < -1 is potentially problematic, as it might indicate lack of adequate variance. This holds true only for the U1 with kurtosis 1.221. However, for practical purposes, this may be problematic only for values > 2.2 or < -2.2 [74]. Hence, U1 is not excluded from further analysis, and all variables are retained.

B. Exploratory Factor Analysis

Factor analysis is used for exploring data patterns, for data reduction, confirming a hypothesis for a factor structure, etc. Exploratory Factor Analysis (EFA) is a multivariate statistical method used to identify the underlying relationships between measured variables [75]. EFA does not discriminate between variables and is considered as an independent technique that does not specify formal hypotheses. Hence, it allows to determine the factors that exist in the dataset and can be used as a cautionary step in order to investigate the match or diversion with the in theory considered factors.

The Kaiser-Meyer-Olkin (KMO) statistic is a Measure of Sampling Adequacy (MSA) and tests whether the partial correlations among variables are small. The calculated value for the dataset is .841 and since it is $> .8$, it is characterized as meritorious [76]. This is an indication that factor analysis will be useful for these variables. A statistically significant result of Bartlett's test of sphericity indicates that the matrix is not an identity matrix and that the variables relate sufficiently each other in order to run a meaningful EFA. Hence, the EFA can proceed and investigate the factors that emerge from the dataset. The Kaiser [77] rule to determine the number of factors is utilized, as it is the most common method used in practice [78].

The EFA was conducted using Maximum Likelihood with Promax rotation in order to assess if the observed variables were adequately correlated and met the criteria of reliability and validity. Maximum Likelihood estimation was chosen in order to determine the unique variance among items and the correlation between factors. In addition, since CFA, and more specifically Structural Equation Modeling (SEM), that will follow after EFA, uses Maximum Likelihood, it was also selected for EFA in order to be consistent. Promax was chosen because it can account for the correlated factors.

Table I
CRONBACH'S α FOR THE FACTORS

	Survey questions	**Cronbach's α**
Utilitarianism	U1,U2,U3,U4	.825
Deontology	D1,D2,D3,D4	.849
Relativism	R1,R2,R3,R4	.957
Absolutism (monism)	A1,A2,A3,A4	.848
Pluralism	P1,P2,P3,P4	.818
Acceptance	AC1,AC2,AC3,AC4	.892

In line with Kaiser's [77] recommendation that only eigenvalues ≥ 1 should be retained, six factors are identified overall, that explain 66.636% of the total variance. This finding is in-line with the number of factors proposed in the theoretical model described in Figure 1 and investigated in this work. There is a clear loading on factors, and the factors demonstrate sufficient convergent validity as their loads are above the minimum recommended threshold of approx. 0.5 [79] for the sample of 126 items used in this research. All factors demonstrate sufficient discriminant validity as no correlations are above 0.700 and there are no problematic cross-loadings. For reliability, the Cronbach's α [52], which is a coefficient of internal consistency, is measured for each factor. As it can be seen in Table I, all of them are $> .8$, which indicates good internal consistency [80].

C. Structural Equation Modeling

As a subsequent step to the Factor Analysis approach followed, CFA, and more specifically SEM, is applied. The SEM model is shown in Figure 2, and is assessed below with the statistics calculated with the IBM AMOS tool that is used to run SEM.

There are several fitness measures to assess a model [81]. The χ^2 (CMIN) divided by the degrees of freedom (DF), leads to the computation of the relative χ^2 (CMIN/DF), which is 1.039. In literature, the relative χ^2 is used as a measure of fitness. It is noted that: " [82] suggest that the researcher also computes a relative chi-square …They suggest a ratio of approximately five or less "as beginning to be reasonable". In our experience, however, χ^2 to degrees of freedom ratios in the range of 2 to 1 or 3 to 1 are indicative of an acceptable fit between the hypothetical model and the sample data" [83]. The calculated $\chi^2 = 1.039$ is seen as adequate since also it is pointed out that "…it seems clear that a ratio > 2.00 represents an inadequate fit" [84].

The Goodness of Fit Index (GFI), and the GFI adjusted for degrees of freedom (AGFI) [85,86] are two others measures of fitness. Both GFI and AGFI should be ≤ 1, where a value of 1 indicates a perfect fit. Both

Preprint version of doi:10.1109/TEM.2018.2877307

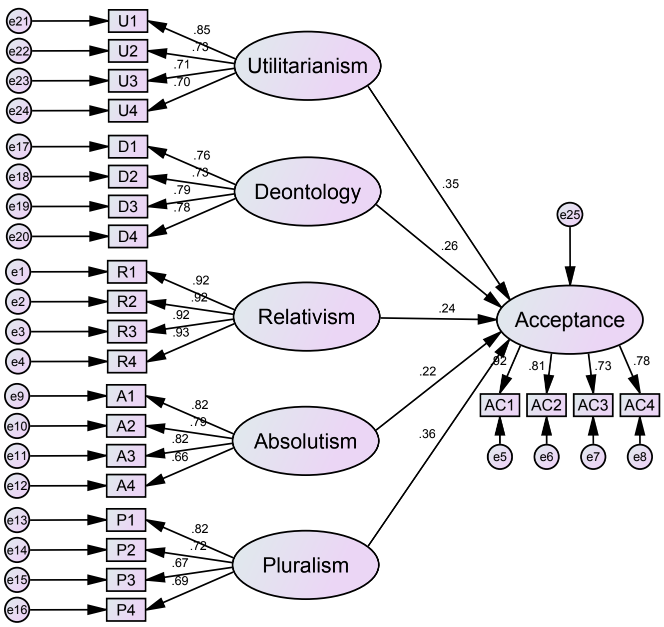

Figure 2. Structural Model in SEM

measures are affected by sample size and therefore the current consensus seems to tend towards not using them [87]. For this model, $GFI = .857$ and $AGFI = .826$ are measured, which are seen as moderate. This might be an effect of the sample size $N = 126$, and as proposed [87] they are not further discussed. Comparative Fit Index (CFI) is another measure [88], and values near to 1 indicate a very good fit. For this model $CFI = .994$ which indicates an excellent fit.

The Root Mean Square Error of Approximation (RMSEA) measures the discrepancy between the fitted model and the covariance matrix in the population. It is noted that "Practical experience has made us feel that a value of the RMSEA of about .05 or less would indicate a close fit of the model in relation to the degrees of freedom" [89]. Values of 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre fit, respectively [90]. The measured $RMSEA = 0.018$ indicates an excellent fit of the model.

The output of the IBM AMOS tool shown in Figure 2, depicts several aspects relevant to the model and represents the result of the hypothesis testing. In the model, all factors are shown with values on the arrows representing the path coefficients (standardized estimates) which show the weight of the links in the path analysis. As it can be seen, all five factors positively contribute to the Acceptance (of self-driving cars).

The Critical Ratio (CR) is the division of the regression weight estimate, by the estimate of its standard error, and tests for loading significance. A CR > 1.96

Table II
TESTING OF HYPOTHESIZED IMPACT ON SELF-DRIVING CAR ACCEPTANCE

Hypothesis	Path	Path Coefficient Weight	CR value >1.96	Support Decision
H1	Utilitarianism → Acceptance	.350	4.061	Supported
H2	Deontology → Acceptance	.265	3.090	Supported
H3	Relativism → Acceptance	.244	3.100	Supported
H4	Absolutism (monism) → Acceptance	.218	2.617	Supported
H5	Pluralism → Acceptance	.362	4.089	Supported

(or < -1.96) indicates two-sided significance at the customary 5% level [91]. Table II shows the path coefficient weight and CR value for all hypotheses. As it can be seen, all of them have a $CR > 1.96$ and therefore all of the hypotheses H1–H5, are supported by the empirical data. Although all hypotheses are in tandem with theory and supported by the empirical data, one has to keep in mind that this does not imply proof [47].

VI. Discussion

Modern self-driving car manufacturers do not yet sufficiently address key issues pertaining ethical aspects in automated decision-making processes of the self-driving cars, although some initial efforts in the research community emerge [17]–[19]. This work makes the link between ethical frameworks and self-driving car acceptance explicit, while also measuring their correlation with statistical methods (SEM). All five posed hypotheses (H1–H5), i.e. that the selected ethical frameworks have an impact on the acceptance of the self-driving cars, are supported by the empirical data collected, as discussed in section V and summarized in Table II. Independent of this affirmation, looking at the discrete answers collected by the survey, additional qualitative insights may be obtained. Some of these are shortly discussed below.

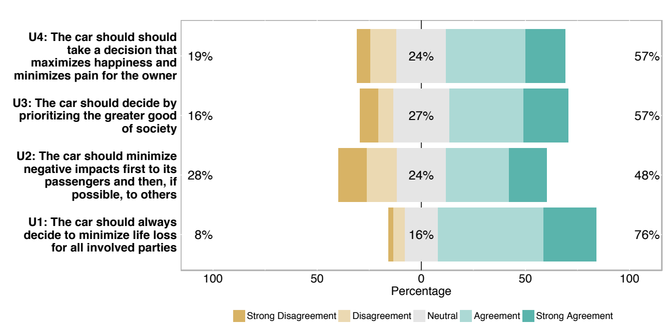

Figure 3. Overview of Utilitarianism responses

A closer look at Utilitarianism results shown in Figure 3 reveals that most people consider that an assessment of some kind ought to be done by the self-driving car and be integrated into its decision algo-

Preprint version of doi:10.1109/TEM.2018.2877307

rithms. Hence aspects such as the greater good of the society, minimization of life loss and negative impacts, etc., ought to be considered by car manufacturers and technology developers. Especially minimization of life-loss is strongly supported in the survey, and it is in-line with the general expectation that self-driving cars will minimize accidents and save human lives, potentially better than human drivers do. Recent research shows that "in the confined scope of unavoidable collisions in road traffic, simple value-of-life models approximate human moral decisions well" [20]. If such direction is taken, considerable thought needs to flow towards how such calculations are done, and avoid potential misuse of it, that may be biased. However, ethical commissions, e.g., in Germany [5], have already spoken against such utilitarian decisions based on personal features such as age, gender, physical or mental constitution.

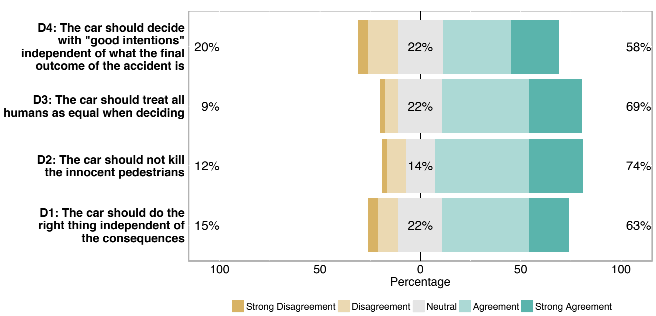

Figure 4. Overview of Deontology responses

Deontology implies that there is an expectation that the self-driving cars carry out their duties with good intentions independent of consequences. As seen in Figure 4, the prevalent view is that cars should treat all people on an equal basis (hence not assigning values to individual people as utilitarianism suggests), as well as trying to protect the innocent pedestrians. This is a strong indication that universal values, especially pertaining human rights [58], ought to be reflected in the car's decision algorithms, and not be subject to selection or adjustment by other stakeholders e.g. the car industry or political forces in each country. However, here also care needs to be taken as deontological positioning is not always understood nor justified.

Absolutism (monism) propagates the existence of global moral values, norms, beliefs, and practices that are praised by the those who agree while they are condemned by those who disagree. Such views propagate group beliefs and may create tensions in society, as shown in the wide-spread of replies in question $A4$ in Figure 5, whether life is sacred and knowingly killing people by a machine would be acceptable. As shown

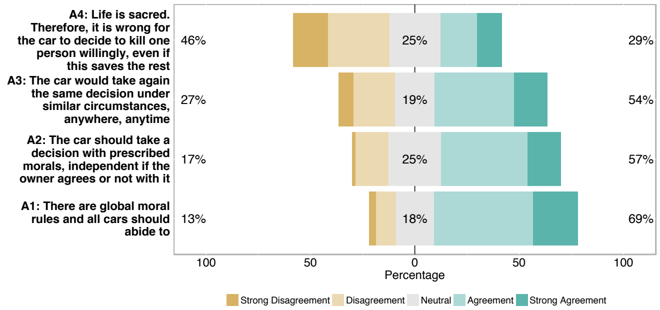

Figure 5. Overview of Absolutism (monism) responses

in Figure 5 there is a strong positioning that the car should have such ethics, and take life and death decisions independently if its owner agrees to it or not. This has several implications, as it would mean that self-driving cars would behave differently than their owners might wish, and raises concerns if cars that do so would actually be bought by people who disagree with their car's decisions in critical situations.

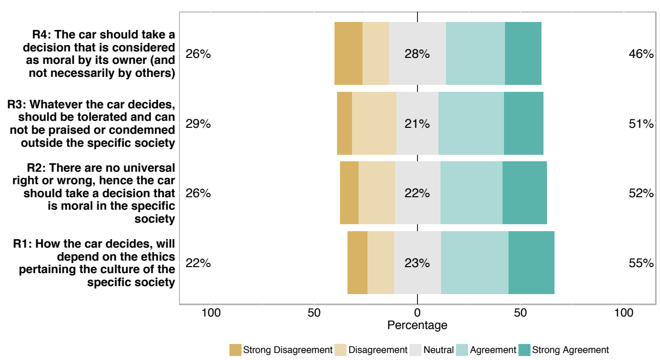

Figure 6. Overview of Relativism responses

Relativism affirms tolerance and is bound to culture, time, society, which may ease the acceptance of decisions taken by self-driving cars in critical situations. As shown in Figure 6, people consider that the self-driving car ought to take into account such ethics in its decisions. Such considerations may reflect the diversity of cultures and philosophies found in the world, but may also create "deadlocks" where specific decisions of the self-driving car, cannot be praised or condemned. If cars reflect the society's heterogeneity, and all occupy the same streets, a question that arises is how such behaviors will be treated, e.g., with respect to liabilities in civil law or justifying car's decisions under criminal law.

Pluralism, propagating the plurality of moral truths, provides a balance among the highly heterogeneous world, tolerance and basic human values such as human rights. Hence, ethical differences may be approached at a global scale. This is also reflected in the views captured

Preprint version of doi:10.1109/TEM.2018.2877307

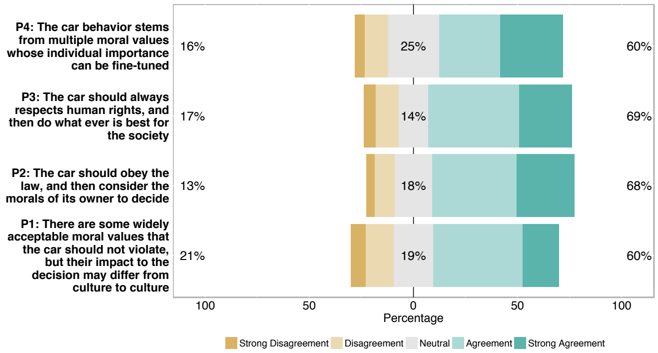

Figure 7. Overview of Pluralism responses

in Figure 7, where a mix of aspects is shown, e.g., the owner's or society's moral views should be considered, while law and global ethical values are ought also to be respected. Therefore, the pluralism framework is seen as a good candidate for decision-making in self-driving cars. However, due to the multiple perspectives that need to be incorporated, it is also a highly complex one, and hence not easy to realize it, especially in contemporary society where "simple" formulas for decisions and "rule of thumb" are sought by ready-made algorithms without really looking at all aspects properly and in a solid manner.

Finally, the survey also measured some aspects of the self-driving car acceptance as shown in Figure 8, from which it is evident that there is a need for ethics to be embedded in self-driving cars. People seem to trust self-driving cars, and therefore they would opt to buy them once they are available, and may prefer them over the normal (non-self-driving) ones. Overall there is a very strong view, that the society needs self-driving cars, as their benefits for a safer and more inclusive society cannot be overseen.

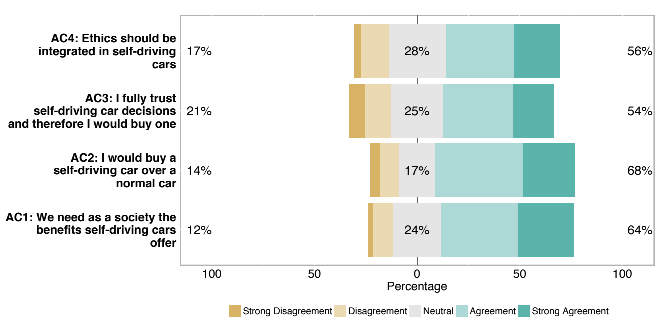

Figure 8. Overview of self-driving car Acceptance responses

The quest to find the role of the ethical frameworks in self-driving car acceptance indicates that our hypotheses were in line with the empirical data and that there is a measurable effect. However, the overall strong support for all frameworks means that there is no clear suggestion, at least from this research, that there should be a preference for a specific framework in the self-driving cars, and no one-size-fits-all solution can be proposed. On the contrary, since all of them seem to have an impact, different parts of the society and people may have different needs and preferences. One thing is clear; that the ethical frameworks considered in this research need to be investigated in-depth, not only qualitatively, but also with mass-scale quantitative surveys as part of the overall research priorities set for AI [92].

Knowledge derived from this research can benefit several stakeholders involved directly and indirectly in the scope of self-driving cars, but also potentially in autonomous and intelligent systems at large [14]. Especially car manufacturers and software engineers developing the car's algorithmic behavior when this pertains to decision-making in critical situations will benefit as they can have some insights on the impact of the ethics supported by the car and their subsequent influence on the car acceptance. The latter will reflect the business success of a specific self-driving car, its market adoption areas, as well as the features/customizations it has to offer with respect to ethics. Policymakers can also benefit [93], as they can consider the outcomes in planned regulations in the area. Individual citizens, both prospective self-driving car buyers as well as the rest, can benefit from the knowledge of the car's behavior, societal needs, and the impact on their community. Finally, the critical discussions on the result may also assist researchers in the multi-disciplinary area of technology, ethics, and law to get new insights and investigate in detail, especially empirically, several of the aspects that are discussed points in this work.

VII. Future Directions

The area of AI that deals with device to device (D2D) communication gains importance as it is strongly coupled with the way the car perceives its environment via its sensors, how it takes decisions based on the internal logic, and how it can communicate its behavior or other info to other parties (e.g. when interacting with infrastructure services or other cars). It is high time to investigate in detail the ethical angle of issues that pertain to the acceptance of self-driving cars, especially from the diverse viewpoints of the multiple stakeholders involved in their lifecycle. As such, an intersectional analysis pertaining law, society, economy, culture, etc. may be the proper way to move forward and tackle issues raised in this work. This by itself is a highly complex and challenging issue, as the expertise required is both in-depth for these domains as well as overall, which

Preprint version of doi:10.1109/TEM.2018.2877307

underlines the need to bring all stakeholders on the table, on the same page and with the willingness to effectively address the tough issues raised.

To exemplify the challenges beyond what has already been discussed in the previous sections, we put some thought-provoking aspects forward, that denote future research directions. For instance, if the self-driving cars roam the roads and autonomously take decisions, would people adjust their behavior? One might assume that they feel safer, and as such, they might not be as careful in their behaviors. For instance, pedestrians may not look out for cars when crossing the street, as they always assume that the car will be taking all necessary measures to protect them. Similarly, the drivers of the cars may not feel ethically responsible for any damages inflicted, as someone else (i.e. the car) takes the decision. Such reliance on automation may be problematic, especially during the transitional phases of the society.

Another challenging issue relates to the ethical conflicts. For instance, assuming that the ethics on the car conflict with the beliefs and ethics of the buyer (or user), would then s/he actually buy or use the car? One way is to set the same ethical framework for all cars, and therefore the customer would have to either accept the fact that this is the case when buying the car (or when using one, e.g., in self-driving taxis). Deviations though to the ethical frameworks supported might lead to segregation of the cars and their behaviors as well as to the people that use them. Regulation is the key here, but even then, there are problematic areas for cross-country differences, as we have already discussed that differences in ethical frameworks might lead to different behaviors reflecting different country laws, culture, etc.

Bias is an area which also needs to be considered as a grand challenge to be addressed. In current self-driving car learning algorithms, there is an implied reward function, that is utilized by the car in order to learn from existing big data. Bias may be hidden in the data, and this may become evident in specific operational situations as an end-effect, which was not possible to identify it at an earlier stage, e.g., design time. Such bias is potentially also linked to ethical frameworks, and as such approaches that investigate it and make it evident from the available data as early as possible are considered critical towards moving ahead. As such, processes and tools need to be developed to detect and rectify such behaviors. Bias, as well as utilitarian approaches, may open the backdoor for misuse. For instance, biased (e.g., racist) behaviors (willing or unwilling) due to errors or manipulation in the decision-making process of the car may emerge. And still, at least for the utilitarian approaches, there is the issue of calculating the utility function as discussed, which for instance may assess the "value" of a pedestrian with non-acceptable criteria e.g., based on ethnicity, age, religion, cultural background, criminal record, ties to the local community, etc. Approaching the issue is challenging, and it is a socio-technical one; hence technology alone can not provide sufficient solutions.

Inherently the trust placed on the discrete processes that are executed, ranging from the information acquired by the car in order to assess a critical situation (e.g. by its sensors or other cars), the flawless execution of the assessment logic, and the guaranteed action once a decision is taken (e.g. brake), raise their own ethical concerns. For instance, another challenge is connected to the potential expectation that all the cars should have the same behavior for the same ethical framework. This implies that the algorithms that implement the specific framework should have the same end-effect for all self-driving cars in similar situations. However, with the different implementations per manufacturer, as well as with the high-variance on the actual capabilities of the car this is expected to be also a challenging issue. For instance, more high-end cars may feature high-performance hardware and software, that may be able to analyze more alternatives and better react to any critical situation than a low-end car. In addition, there is still the question of how long should the car analyze all possible alternatives, and how far into the future, consequences ought to be considered, before taking a decision. Basically, that leads to a new discussion that has practical implications as, e.g., how many solutions in the set problem space need to be evaluated before the car acts on a decision.

A key issue is also linked to trust in the whole lifecycle of D2D processes e.g. information acquisition, interaction with other cars, etc. and its potential misuse including bias. While the cars will rely on their sensors, algorithms and ethical frameworks, to take their decisions, malfunctioning or malicious manipulation of any part of this process may lead to decisions that are not compliant with the ethical framework of the car. As an example, some early efforts in self-driving cars would falsely classify both a plastic bag and a rock as similarly dangerous and drive around them, something that no human would do, and which may lead to a critical situation, e.g., causing an accident, rather than just driving over the plastic bag. Fooling the sensor with false data or the image recognition with purposefully prepared images (e.g., adversarial attacks), is a whole new area of attacks that can be performed. As also the city and road infrastructure becomes more complex, the cars will have to rely also on infrastructure (even

Preprint version of doi:10.1109/TEM.2018.2877307

Internet-based) services (to acquire location-specific information), and potentially to the interaction with other cars. The trust imposed to these services as well as their impact in the decision making processes of the car is another key challenge to address. Furthermore, privacy is another satellite issue here, as to the extent of the information richness that gets exchanged, monitored and processed in D2D scenarios, in order to realize the advanced functionalities envisioned in the self-driving cars. All of these increase the target space and bring even more forward the safety, security, trust, privacy and dependability challenges that need to be addressed.

The issue of liability poses another key challenge. As already discussed there is no clarity on who would be responsible for the actions taken in a critical situation, e.g., if the manufacturer or the programmers would be responsible for lives lost due to coding errors, bugs, or biased data used for training the AI in the car. In addition, if each car owner can select the ethical mode of its car, would it actually be fair to put the liability issue on the owner even if s/he has no idea on the actual algorithm that led to the specific decision? To further perplex the issue, there have been discussions about recognizing AI at large (and by extension probably the future self-driving cars) as a legal entity, which brings new challenges for the multitude of aspects pertaining civil and criminal law.

Another promising but challenging investigation direction is the relevance of the population of the self-driving cars with respect to their coordinated capabilities, e.g., as a swarm. Up to now, most research has considered these cars as singletons that act alone in an environment where potentially legacy (non-intelligent) cars are the majority. As such their actions are limited to what they can do by themselves (solely relying on their sensors and logic). However, with an increase in their population, cooperative scenarios gain importance. In contrast to their human users who are unable to cooperate in the context of unavoidable accidents due to time or stress constraints, the self-driving cars could interact with each other (and infrastructure services) and cooperate towards, e.g., preventive measures that would minimize the possibility that an accident happens in the first place. This cooperation opens the capabilities for new actions to be undertaken and enhance the driving safety. However, such scenarios are hardly investigated from the technological, but more importantly from their ethical and regulatory angles. And there are also additional challenges, e.g. if two cars with different ethical frameworks need to cooperate, would they have to negotiate on the outcome and reach a consensus? In that case, would that final collective decision be in conflict with the ethical framework of some of the participating cars in the negotiation? Such questions have multiple dimensions and need to be investigated in order to assess their applicability and implications.

Finally, the side-effects to society at large need to be also investigated, and this is not straightforward. While the fears and effects of roboticization have been argued over decades, a rapid large scale uptake of a disruptive technology such as the self-driving cars may have unexpected impacts in a plethora of everyday aspects. For instance, it is beneficial to society that the number of accidents will be reduced. However, such a reduction will also lead to a shortage of organs available for transplant. As such, it becomes evident that efforts towards engineering artificial organs ought to be strengthened in the near future in order to address this issue. Hence, the near-future research priorities might be affected, and these have also their ethical dimensions.

VIII. Conclusions

In the near future, millions of self-driving cars are expected to roam the streets and take instant decisions in critical situations that involve life and death of humans. Such machine-driven decision-making process has an ethical dimension, which although not new, is not sufficiently investigated with respect to its impact on the self-driving car acceptance. This work addressed the specific white spot, i.e., if and what impact do the different ethical frameworks have in the self-driving car acceptance.

Five ethical frameworks (Utilitarianism, Deontology, Relativism, Absolutism, and Pluralism) are identified in the literature and are adopted as factors that it was hypothesized could impact self-driving car acceptance. A model investigating their impact was proposed, and a survey was carried out to gather empirical data to test the hypotheses. The results revealed that all five factors have a statistically significant effect on self-driving car acceptance (in the empirical data collected). As such the respective posed hypotheses are supported.

The implications of the findings and the critical discussions in this work, make it evident that more emphasis should be put towards research in this domain, as the success or failure of the market introduction of self-driving cars could be affected by the ethical frameworks they employ in order to derive their decisions. Hence, all stakeholders, e.g., the technology providers, the legislators, the consumer associations, and the manufacturers, ought to consider the discussed insights as food for thought, in order to further investigate in detail the role of ethics in self-driving car critical decision-making, prior to their eminent public introduction. It has also become

Preprint version of doi:10.1109/TEM.2018.2877307

evident from the discussions and challenges presented, that the issue addressed in this work is part of a huge area with many dimensions, and as such it is complex by itself with many inter-twined aspects that need to be carefully addressed in an interdisciplinary manner.

References

[1] D. Begg, "A 2050 vision for London: What are the implications of driverless transport," Tech. Rep., 2014. [Online]. Available: http://www.transporttimes.co.uk/Admin/uploads/64165-transport-times_a-2050-vision-for-london_aw-web-ready.pdf

[2] J. Greenough, "10 million self-driving cars will be on the road by 2020," Jun. 2016, business Insider. [Online]. Available: https://goo.gl/e4DbJe

[3] C. Urmson and W. R. Whittaker, "Self-driving cars and the urban challenge," *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 66–68, Mar. 2008.

[4] P. Valdes-Dapena, "Volvo promises deathproof cars by 2020," Jan. 2016. [Online]. Available: http://money.cnn.com/2016/01/20/luxury/volvo-no-death-crash-cars-2020

[5] Ethik-Kommission, "Automatisiertes und Vernetztes Fahren," Tech. Rep., Jun. 2017. [Online]. Available: http://www.bmvi.de/bericht-ethikkommission

[6] NHTSA, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," Tech. Rep., 2015. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

[7] NTSB, "Preliminary report highway HWY18MH010," Tech. Rep., 2018. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

[8] N. Goodall, "Ethical decision making during automated vehicle crashes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2424, pp. 58–65, Dec. 2014.

[9] P. Carsten, T. R. Andel, M. Yampolskiy, and J. T. McDonald, "In-vehicle networks: Attacks, vulnerabilities, and proposed solutions," in *Proceedings of the 10th Annual Cyber and Information Security Research Conference on - CISR '15*, 2015.

[10] J. J. Thomson, "The trolley problem," *The Yale Law Journal*, vol. 94, no. 6, p. 1395, May 1985.

[11] N. J. Goodall, "Can you program ethics into a self-driving car?" *IEEE Spectrum*, vol. 53, no. 6, pp. 28–58, Jun. 2016.

[12] J. Bryson and A. Winfield, "Standardizing ethical design for artificial intelligence and autonomous systems," *Computer*, vol. 50, no. 5, pp. 116–119, May 2017.

[13] S. Karnouskos, "Ethical frameworks in critical decisions and the acceptance of self-driving cars," Master's thesis, Stockholm University, 2017.

[14] "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," resreport, 2018. [Online]. Available: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

[15] European Parliament, "Ethical Aspects of Cyber-Physical Systems," Scientific foresight study,, Jun. 2016. [Online]. Available: https://goo.gl/Fp5Wjs

[16] P. Lin, "Why ethics matters for autonomous cars," in *Autonomes Fahren*. Springer Nature, 2015, pp. 69–85.

[17] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science*, vol. 352, no. 6293, pp. 1573–1576, Jun. 2016.

[18] M. Sikkenk and J. Terken, "Rules of conduct for autonomous vehicles," in *7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*, Jun. 2015.

[19] A.-K. Frison, P. Wintersberger, and A. Riener, "First person trolley problem," in *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - Automotive'UI 16*. Association for Computing Machinery (ACM), 2016.

[20] L. R. Sütfeld, R. Gast, P. König, and G. Pipa, "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*, vol. 11, Jul. 2017.

[21] C. Ess, *Digital Media Ethics*, 2nd ed., ser. Digital Media and Society. Polity Press, 2014.

[22] World Economic Forum, "Self-driving vehicles in an urban context," Nov. 2015, press Briefing. [Online]. Available: https://goo.gl/IPVVtf

[23] A. Hevelke and J. Nida-Rümelin, "Responsibility for crashes of autonomous vehicles: An ethical analysis," *Science and Engineering Ethics*, vol. 21, no. 3, pp. 619–630, Jun. 2014.

[24] M. A. Nees, "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*, vol. 60, no. 1, pp. 1449–1453, Sep. 2016.

[25] M. Kyriakidis, R. Happee, and J. de Winter, "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 32, pp. 127–140, Jul. 2015.

[26] Accenture, "Consumers in us and uk frustrated with intelligent devices that frequently crash or freeze, new accenture survey finds," Feb. 2011. [Online]. Available: https://newsroom.accenture.com/article_display.cfm?article_id=5146

[27] L. Yvkoff, "Many car buyers show interest in autonomous car tech," Apr. 2012. [Online]. Available: https://www.cnet.com/roadshow/news/many-car-buyers-show-interest-in-autonomous-car-tech/

[28] B. Schoettle and M. Sivak, "A survey of public opinion about autonomous and self-driving vehicles in the u.s., the u.k., and australia," Tech. Rep., 2014. [Online]. Available: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/108384/103024.pdf

[29] CISCO, "Consumers desire more automated automobiles, according to cisco study," May 2013. [Online]. Available: https://newsroom.cisco.com/press-release-content?articleId=1184392

[30] C. Hohenberger, M. Spörrle, and I. M. Welpe, *How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups*. Elsevier BV, Dec. 2016, vol. 94.

[31] PWC, "Driving the future: understanding the new automotive consumer," Tech. Rep., 2016. [Online]. Available: http://www.pwc.com/us/en/industry/entertainment-media/publications/consumer-intelligence-series/assets/pwc-autotech-v18.pdf

[32] P. Bansal, K. M. Kockelman, and A. Singh, "Assessing public opinions of and interest in new vehicle technologies: An austin perspective," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 1–14, Jun. 2016.

[33] S. V. Casley, A. S. Jardim, and A. M. Quartulli, "A study of public acceptance of autonomous cars," BSc. Thesis,, Worcester Polytechnic Institute, 2013. [Online]. Available: https://web.wpi.edu/Pubs/E-project/Available/E-project-043013-155601/unrestricted/A_Study_of_Public_Acceptance_of_Autonomous_Cars.pdf

Preprint version of doi:10.1109/TEM.2018.2877307

[34] W. Payre, J. Cestac, and P. Delhomme, "Intention to use a fully automated car: Attitudes and a priori acceptability," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 252–263, Nov. 2014.

[35] P. Bansal and K. M. Kockelman, "Are we ready to embrace connected and self-driving vehicles? a case study of texans," *Transportation*, Nov. 2016.

[36] D. Howard and D. Dai, "Public perceptions of self-driving cars: The case of berkeley, california," Tech. Rep., 2013, 93rd Annual Meeting of the Transportation Research Board. [Online]. Available: http://www.danielledai.com/academic/howard-dai-selfdrivingcars.pdf

[37] IpsosMORI, "Only 18 per cent of britons believe driverless cars to be an important development for the car industry to focus on," Jul. 2014, ipsos MORI Loyalty Automotive Survey. [Online]. Available: https://goo.gl/S7TZa8

[38] J. Youngs, "2014 u.s. automotive emerging technologies study results," May 2014. [Online]. Available: http://www.jdpower.com/cars/articles/jd-power-studies/2014-us-automotive-emerging-technologies-study-results

[39] Continental, "Continental mobility study 2013: Findings to the acceptance of advanced driver assistance systems and automated driving," 2013. [Online]. Available: https://goo.gl/JbVGGa

[40] C. J. Haboucha, R. Ishaq, and Y. Shiftan, "User preferences regarding autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 37–49, May 2017.

[41] J. Zmud, I. N. Sener, and J. Wagner, "Consumer acceptance and travel behavior impacts of automated vehicles," Tech. Rep., 2016, pRC 15-49 F. [Online]. Available: http://tti.tamu.edu/documents/PRC-15-49-F.pdf

[42] T. Leicht, A. Chtourou, and K. B. Youssef, "Consumer innovativeness and intentioned autonomous car adoption," *The Journal of High Technology Management Research*, May 2018.

[43] K. Kaur and G. Rampersad, "Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars," *Journal of Engineering and Technology Management*, vol. 48, pp. 87–96, Apr. 2018.

[44] L. M. Hulse, H. Xie, and E. R. Galea, "Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age," *Safety Science*, vol. 102, pp. 1–13, Feb. 2018.

[45] M. Körber, E. Baseler, and K. Bengler, "Introduction matters: Manipulating trust in automation and reliance in automated driving," *Applied Ergonomics*, vol. 66, pp. 18–31, Jan. 2018.

[46] M. D. Myers, "Qualitative research in information systems," Originally published in MISQ Discovery, Living Version., University of Auckland, New Zealand, Jun. 1997. [Online]. Available: http://www.qual.auckland.ac.nz

[47] D. Straub, D. Gefen, and M.-C. Boudreau, "The ISWorld quantitative, positivist research methods website," 2004. [Online]. Available: https://web.archive.org/web/20140819221204/http://dstraub.cis.gsu.edu:88/quant/

[48] D. R. Forsyth, "A taxonomy of ethical ideologies." *Journal of Personality and Social Psychology*, vol. 39, no. 1, pp. 175–184, 1980.

[49] M. Denscombe, *The Good Research Guide for small-scale social research projects*, 4th ed. McGraw Hill, 2010.

[50] B. M. Byrne, *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, Second Edition*, 2nd ed., ser. Multivariate Applications Series. Routledge, Taylor & Francis Group, 2009.

[51] D. Straub, M.-C. Boudreau, and D. Gefen, "Validation guidelines for IS positivist research," *Communications of the Association for Information Systems*, vol. 13, no. 24, pp. 380–427, 2004.

[52] l. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[53] R. K. Yin, *Case Study Research: Design and Methods*, 4th ed. SAGE Publications, Inc, Dec. 2009.

[54] F. Eisenführ, M. Weber, and T. Langer, *Rational Decision Making*. Springer, 2010.

[55] J. C. Gerdes and S. M. Thornton, "Implementable ethics for autonomous vehicles," in *Autonomes Fahren*. Springer Nature, 2015, pp. 87–102.

[56] B. F. Malle, M. Scheutz, T. Arnold, J. Voiklis, and C. Cusimano, "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, 2015.

[57] I. Kant, *Critique of Practical Reason*. Bobbs-Merrill, 1785, translation by Lewis White Beck 1959.

[58] I. Coca-Vila, "Self-driving cars in dilemmatic situations: An approach based on the theory of justification in criminal law," *Criminal Law and Philosophy*, Jan. 2017.

[59] J. Millar, "Technology as moral proxy: Autonomy and paternalism by design," in *Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium on*, May 2014, pp. 1–7.

[60] M. Tedre, "What should be automated?: The fundamental question underlying human-centered computing," in *Proceedings of the 1st ACM international workshop on Human-centered multimedia - HCM '06*. ACM, 2006.

[61] F. Flemisch, M. Heesen, T. Hesse, J. Kelsch, A. Schieben, and J. Beller, "Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations," *Cognition, Technology & Work*, vol. 14, no. 1, pp. 3–18, Nov. 2011.

[62] F. S. de Sio, "Killing by autonomous vehicles and the legal doctrine of necessity," *Ethical Theory and Moral Practice*, Feb. 2017.

[63] J. Borenstein, J. Herkert, and K. Miller, "Self-driving cars: Ethical responsibilities of design engineers," *IEEE Technology and Society Magazine*, vol. 36, no. 2, pp. 67–75, Jun. 2017.

[64] J. Gogoll and J. F. Müller, "Autonomous cars: In favor of a mandatory ethics setting," *Science and Engineering Ethics*, Jul. 2016.

[65] C. Allen, W. Wallach, and I. Smit, "Why machine ethics?" *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 12–17, Jul. 2006.

[66] S. Applin, A. Riener, and M. Fischer, "Extending driver-vehicle interface research into the mobile device commons : Transitioning to (nondriving) passengers and their vehicles." *IEEE Consumer Electronics Magazine*, vol. 4, no. 4, pp. 101–106, Oct. 2015.

[67] N. McBride, "The ethics of driverless cars," *SIGCAS Comput. Soc.*, vol. 45, no. 3, pp. 179–184, Jan. 2016.

[68] J. Petit and S. Shladover, "Potential cyberattacks on automated vehicles," *IEEE Trans. Intell. Transport. Syst.*, pp. 1–11, 2014.

[69] S. Karnouskos and F. Kerschbaum, "Privacy and integrity considerations in hyperconnected autonomous vehicles," *Proceedings of the IEEE*, vol. 106, no. 1, pp. 160–170, Jan. 2018.

[70] K. Ali Alheeti, A. Gruebler, and K. McDonald-Maier, "An intrusion detection system against malicious attacks on the communication network of driverless cars," in *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, Jan. 2015.

[71] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[72] M. Schellekens, "Self-driving cars and the chilling effect of liability law," *Computer Law & Security Review*, vol. 31, no. 4, pp. 506–517, 2015.

Preprint version of doi:10.1109/TEM.2018.2877307

[73] G. Norman, "Likert scales, levels of measurement and the "laws" of statistics," *Advances in health sciences education : theory and practice*, vol. 15, no. 5, pp. 625–632, Feb. 2010.
[74] V. A. Sposito, M. L. Hand, and B. Skarpness, "On the efficiency of using the sample kurtosis in selecting optimal L_p estimators," *Communications in Statistics - Simulation and Computation*, vol. 12, no. 3, pp. 265–272, Jan. 1983.
[75] M. Norris and L. Lecavalier, "Evaluating the use of exploratory factor analysis in developmental disability psychological research," *Journal of Autism and Developmental Disorders*, vol. 40, no. 1, pp. 8–20, 2010.
[76] H. F. Kaiser and J. Rice, "Little Jiffy, Mark IV," *Educational and Psychological Measurement*, vol. 34, no. 1, pp. 111–117, Apr. 1974.
[77] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141–151, Apr. 1960.
[78] L. R. Fabrigar, R. C. Maccallum, D. T. Wegener, E. J. Strahan, R. R. Fabrigar, E. J. Strahan, and D. Of, "Evaluating the use of exploratory factor analysis in psychological research," *Psychological Methods*, pp. 272–299, 1999.
[79] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. Prentice Hall, 2010.
[80] D. George and P. Mallery, *IBM SPSS Statistics 21 Step by Step: A Simple Guide and Reference*, 13th ed. Pearson, 2013.
[81] J. S. Tanaka, "Multifaceted conceptions of fit in structural equation models," in *Testing Structural Equation Models*. SAGE, 1993.
[82] B. Wheaton, B. Muthén, D. Alwin, and G. Summers, "Assessing reliability and stability in panel models." *Sociological methodology, San Francisco: Jossey-Bass*, pp. 84–136, 1977.
[83] E. Carmines and J. McIver, "Analyzing models with unobserved variables," *Social measurement: Current issues, Beverly Hills: Sage*, 1981.
[84] B. M. Byrne, *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer-Verlag, 1989.
[85] K. Jöreskog and D. Sörbom, *LISREL-VI: user's guide; analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*, 4th ed. Mooresville, Ind.: Scientific Software, 1986.
[86] J. Tanaka and G. Huba, "A fit index for covariance structure models under arbitrary gls estimation," *British Journal of Mathematical and Statistical Psychology*, vol. 38, pp. 197–201, 1985.
[87] S. Sharma, S. Mukherjee, A. Kumar, and W. R. Dillon, "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models," *Journal of Business Research*, vol. 58, no. 7, pp. 935–943, 2005.
[88] P. Bentler, "Comparative fit indexes in structural models," *Psychological Bulletin*, vol. 107, pp. 238–246, 1990.
[89] M. W. Browne and R. Cudeck, *Alternative ways of assessing model fit*. Newbury Park, CA: Sage, 1993, pp. 136–162.
[90] R. C. MacCallum, M. W. Browne, and H. M. Sugawara, "Power analysis and determination of sample size for covariance structure modeling." *Psychological Methods*, vol. 1, no. 2, pp. 130–149, 1996.
[91] J. Hox and T. Bechger, "An introduction to structural equation modeling," *Family Science Review*, vol. 11, pp. 354–373, 1998.
[92] S. Russell, D. Dewey, and M. Tegmark, "Research priorities for robust and beneficial artificial intelligence," *AI Magazine*, vol. 36, no. 4, 2015.
[93] S. Li, P.-C. Sui, J. Xiao, and R. Chahine, "Policy formulation for highly automated vehicles: Emerging importance, research frontiers and insights," *Transportation Research Part A: Policy and Practice*, May 2018.

Preprint version of doi:10.1109/TEM.2018.2877307