A Moral Consensus Mechanism for Autonomous Driving: Towards a Law-compliant Basis of Logic Programming

Jonas Wanner[1], Lukas-Valentin Herm[1], Marvin Langer[1], Florian Imgrund[1], and Christian Janiesch[1]

[1] Julius-Maximilians-Universität Würzburg, Würzburg, Germany
{name.surname}@uni-wuerzburg.de, langer.marvin@gmail.com

Abstract. Research into autonomous vehicles is making progress. While implementation is progressing through machine learning and efficient sensor technology, one key challenge remains dealing with moral disputes. In general, traffic requires for moral decisions that might even decide on the life or death of participants. While people make intuitive decisions in accidents, a decision of an autonomous vehicle is made already at the programming stage. Thus, a concrete handling for implementation is needed. Due to a lack of legislation, this is still missing and prevents car manufacturers from a practical solution. The paper at hand addresses this problem by presenting a consensus mechanism, combining moral convictions, legislation, and programming guidelines. Based on a study of dilemma situations, moral principles of the 'correct action' of autonomous vehicles are derived. Of four principles, we confirm one, reject two, and propose one for further research investigation to form a basis for jurisdictions.

Keywords: Autonomous Driving, Dilemma Situations, AI Ethics, Moral Machine, AI Programming Guidelines

1 Introduction

Germany is struggling for its global top position in the automotive sector. Recently, it seems to lose the battle for key innovations on data-driven services and autonomous driving [1]. The latter is regarded as the next top disruptive innovation in automobility in the upcoming years – and will have a major impact on society [2]. The diffusion of the technology requires not only user acceptance, but also legal certainty for automobile manufacturers to implement their artificial intelligence solution. This hampers the development process and a practical application [3]. Yet, in the event of an accident by an autonomous vehicle, manufacturers can be sued for damages to an uncertain extent. For example, in 2016, when the first person was killed by an autonomous vehicle of Uber, the firm entered into a public lawsuit with an existing gap on legal clarity [4].

In order to achieve practical suitability, the question of the 'correct action' of an autonomous vehicle must be declared. This applies in particular to dilemma situations

15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany

https://doi.org/10.30844/wi_2020_a2-wanner

(DS). A DS is an extreme situation in which there is no way out between the decision of life and death of several traffic participants. In contrast to the unconscious (manual) decision of man, a machine's decision is already determined in its programming stage, and thus ex-ante. This makes a law-compliant basis for programming logic inevitable. Otherwise, it will impede manufacturer's development and technological diffusion [2].

Germany's government already recognized the problem. Together within members of the UN, an amendment to the Vienna Convention on Road Traffic was achieved in 2016. Subsequently, Germany adopted its first legal enactment in 2017. It allows for fully automated, but not for autonomous driving [5]. The driver must still continue to participate attentively in traffic. Following an exchange with domain experts, an extension of the legislation seems to have failed primarily due to a lack of moral concepts and technical understanding. While technical investigations are often examined by expert groups and research institutions, moral investigations are more difficult. An innovative, unconventional approach aims to fill these gaps.

While there is research on the technical appearance and design of autonomous agents, see [6] for a summary, our study is set in a separate, more immature field of research. Expected reaction of autonomous vehicles in DS are tested to reveal moral similarities. Its major contribution is the one of Awad et al. [7] from 2018. The team carried out an extensive global study on the 'correct action' of an autonomous vehicle. Participants were asked to answer various, pre-defined DS scenarios. Any uncertainties of the outcome were excluded to enable the identification of moral similarities. As this type of study design was never used before within a legislative process it could be an opportunity for the German government to catch up with more innovative countries in autonomous driving. The assumed suitability is derived from the previous basis for new law-making in Germany: The (aim for the) use of moral similarities.

The paper at hand addresses this issue by developing a proof of concept (PoC) [8] mechanism for a moral consensus derivation on autonomous driving. By the nature of PoC, we strive for the fundamental feasibility of our approach. In contrast to existing research, the approach is differentiated by its first integration into an overarching legislative process. Our research artefact is based on a structured literature review, an expert survey, and a subsequent web-based survey on Moral Machine DS scenarios. In summary, we define the following research question (RQ):

RQ: *How can we extend Germany's law-making process by the derivation of moral concepts for the topic of autonomous driving in order to facilitate the overall process?*

We apply the Design Science Research (DSR) approach to answer our stated research question. Therefore, the paper is structured as follows: In Section 2, we present the theoretical basis on autonomous driving and related topics regarding decision-making in DS. Section 3 comprises the description of the design of the DSR approach used and the surveys conducted. Section 4 demonstrates the development of the moral consensus mechanism by literature and expert statements, ending with the

https://doi.org/10.30844/wi_2020_a2-wanner

final conceptualization. Further, we evaluate our artefact with real-world data. In Section 5 we discuss our research contribution and related limitations. Finally, in Section 6 we summarize our findings and give an outlook.

2 Theoretical Foundation

2.1 Autonomous Driving

Autonomous driving (AD) is defined as the self-determined action of an intelligent vehicle within the limits of specified norms [9]. Today, there is a distinction between different types of AD due to the degree of mechanical performance in combination with the demands on the occupants during use of the vehicle. A total autonomous vehicle is fully automated and represents the highest level of automation [10].

The division of the Society of Automotive Engineers (SAE) separates six types of degrees of autonomous vehicles on a scale from 0 to 5. This corresponds to other institutions such as the German Association of the Automotive Industry (VDA) [11], starting from (0) no automation, without any driver assistance, through (5) full automation, without a driver necessary at all [12]. However, today's stage of development in Germany is still on the preliminary stages of (2) partially to (4) highly automated driving [11].

2.2 Dilemma Situations

Dilemma situations exist in every stage of AD and occur in every level. A DS is an extreme situation in which the driver has no other choice expect of the decision for or against the life and death of specific traffic participants. In non-AD, it is the driver's responsibility to choose the right course of action within a fraction of seconds. So, it is not possible to weight the situation in detail. The driver will always act subconsciously. The same is happening in level five of autonomous driving, except that the driving algorithm decides. Thus, algorithms should be developed based on ethical theories and existing laws to prevent unequal treatment of groups of people [13]. However, theses algorithms have immense upsides. While the human driver has to act immediately in a DS and therefore often decides subjectively without ethical reflection, a pre-defined behavior for the algorithm based on comprehensively evaluated ethical and moral principles can be achieved [14].

In order to derive the 'correct action' in such a situation, a so-called 'trolley problem' research design shall shed light on this. It extends the explained dilemma by a theoretical moral experiment. A perfect information situation is assumed without any uncertainties of the outcome. Also, technical conditions are not considered to distinguishing features from complex facts. The aim is to isolate the moral criteria [15]. Here, the respondent must choose from two known event possibilities in each case in a series of DS scenarios depicted [7]. The results can be moral similarities transferred into moral principles for programming guidelines.

https://doi.org/10.30844/wi_2020_a2-wanner

2.3 Ethical Theories

Ethics seeks answers to questions of human coexistence in which previous ways of life lose their validity. Thereby, norms are formulated as accepted guidelines for individual behavior based on values, which themselves are denote as the common perception of reality. The understanding of common values ensures an effective community [16]. It forms a reciprocal state of restriction and possibility. In exemplary pedestrians at a crosswalk can pass the road safely, as a driver of an arriving car will stop. The totality of the prevailing values and norms is called morality. In addition, ethics refers to the theoretical discourse of moral questions [17].

Numerous ethical approaches and theories exist. They compete in value concepts and related explanations, especially in ethical dilemmas [18], for example given in DS. For deontologists, good will is crucial and not the result of an action. Thus, the moral goodness of a decision lies in its intentions. The consequences of an action are not decisive. So, even if the negative consequences predominate, the underlying act can be considered correct [17]. This forms actual German legislative for the jurisdiction in DS. If a human driver is confronted with a DS and has to decide within fractions of a second, he will not be guilty for the predominant negative consequences as he was not able to act consciously. This contradicts the teleological ethics. Here, consequences are the central point of a decision. By weighing them, it can be decided whether an action is right or wrong. In the related utilitarianism approach [19] the focus is on the 'greatest happiness of the greatest number' [20].

2.4 Laws by Moral Similarities

Moral similarities can also be found directly and indirectly in laws as the passenger's right of a crosswalk in German's StVO§26. It can even shape the legislation process, as, for example the instance of automated driving by the state of California (US). This is regulated in California Vehicle Code (CVC) Section 38750 as well as Code of Regulation §227 at California DMV, which defines not only autonomous vehicles, necessary control mechanisms by the driver and technology, but also the liability for accidents in autonomous operation by the manufacturer and a necessary insurance. There is no precise legal consideration of problems that may arise in the context of autonomous driving, for example moral problems. Similarly, no precise rules for the compliant development by the vehicle manufacturers are specified in this context. Thereby the programmer has to adhere to ethical moral concepts. Such a law definition process based on a moral consensus for AD is even conceivable in Germany, which is missing so far. This is proven, for example by the revision of the existing stem cell law (StZG) of the Federal Republic of Germany [21, 22]. Hence, §1 of the StZG also reflects moral concepts such as 'human dignity' [23].

https://doi.org/10.30844/wi_2020_a2-wanner

3 Research Design

3.1 Overall Methodology

For the development of our scientific artefact we apply the design science research approach according to Peffers et al. [21]. DSR is as an iterative and stepwise research development method. It comprises six development stages: *problem identification & motivation*, *definition of objectives of a solution*, *design & development*, *demonstration*, *evaluation*, and *communication* [21].

As a general problem, we have noticed that in both the German and the international context there is no established standardized approach for an AD law-making process. Especially in countries such as Germany, the prior definition of laws is necessary to enable integration into road traffic. Both a technical understanding and moral foundations are necessary to develop new laws for AD. Maintaining to the example of Germany, we aim to address this issue by developing a moral consensus mechanism that supports and accelerates the legislative process.

In the design and development stage, we conducted three iterations. We began with a systematic literature analysis, based on the recommendations of Webster and Watson [22]. Afterwards, there was an (re-)evaluation of the theoretical-based construct by experts interviewed. This was followed by an (re-)modification.

During the demonstration phase, we formalize our moral consensus mechanism. Further, we specified the required subsections and describe each underlying step of our proposed mechanism. The artefact was then evaluated in a practical manner. We use literature, experts, and a web-based survey to do so. Finally, we discuss the principles derived and the applicability of our research artefact.

3.2 Survey Methodology

Expert survey. We employed an expert study. This allowed for a deeper understanding of the topic's interrelations. Furthermore, an evaluation of the research artefact became possible. We designed a semi-structured survey according to [23]. The questionnaire was three-folded: (A) demographics, (B) questions on legislative procedures, and (C) evaluation of the derived research artefact. As interview partners, six legal experts of autonomous driving or related fields of research (e.g. robot law-making) were selected. Two of them were asked twice to validate the modifications. A total of five interviewees came from Germany, and one form the USA. All experts have had at least five years of expertise in the field of interest. The conversation was either by phone or face-to-face. Each interview took between 78 and 102 minutes.

Web-based survey. The derivation of moral similarities was conducted by a Web-based survey. To ensures a unified cultural background [7] we limited our questionnaire to a German version and disseminated through various social media channels. The questionnaire itself was three-folded: (A) demographics, (B) questions on autonomous driving, and (C) 17 different DS scenarios according to the design of

https://doi.org/10.30844/wi_2020_a2-wanner

Moral Machine[1]. We extended the original choice of 'right' and 'left' with 'random'. Thus, we prevent pure utilitarianism in the participant's decision making by weighing lives. With this we also permit for a deontological decision making as the actual way of judgment in German's legislation (cf. Section 2.3). We received a total of 119 valid responses from 19 to 21 October 2018. 51 responses have been excluded due to either incompleteness or a disagreement on build-in control questions following [24]. 60 % of the participants were 20-30 years old and 28.4 % younger than 20 years. 76.1 % of the participants stated that they are pupils, students, or still in education. Further, 87.8 % of the surveyed have at least the university entrance qualification. The participation of male and female participants was balanced.

4 Moral Consensus Mechanism

4.1 State-of-the-Art

We conducted a structured literature review based on the recommendations of Webster and Watson [22]. In order to comprise the interdisciplinary nature of the topic, we selected six databases: IEEE Xplore, AISeL, and ACM Digital Library to cover IT-related engineering-related research areas, Business Source Premier (EBSCO) for economics and business related content, and Web of Science and ScienceDirect for a comprehensive selection of interdisciplinary publications.

We had initially considered limiting the results regarding their (journal) ranking. We discarded this due to the novelty of the topic. Our search term we have used was the following pseudocode: *((autonomous driv* | automated vehicles| driverless car* | self-driving car* | artificial intelligence) AND (consensus | ethic* | moral* | responsibility | legi* | programming | dilemma situation* | liability)).* By extending a forward and backward search with the help of Web of Science, as recommended by [22], we found a total of 1,637 publications. After a full-text analysis, focusing on information about the integration of moral (or ethical) problems into an AD law-making process, 105 contributions remained. We classified them due to their focus on regulatory needs or problems. Both clusters are described below.

Regulatory needs. The need for legal regulation of AD has been confirmed in various research articles, e.g., Cunneen et al. [25] and Baumann et al. [26]. Both stated that there cannot be a general regulation for AI-based development as it will be always related to its application context. Thus, there must be a specific one for AD. Further, a recent study by Schulek-Leech et al. [33] revealed the lack of guidelines, laws, and frameworks. This is absolutely necessary as developers are not incapable to consider accepted ethical or social implications in their algorithms for an unbiased and consistent development. This is already attempted to address but is still in one's infancy. For example, the research of Li et al. [27] defines three groups of interest to be aware of if developing an AD focused framework: automotive manufacturers,

[1] Cf. digital appendix at DOI: 10.13140/RG.2.2.29921.04964

https://doi.org/10.30844/wi_2020_a2-wanner

consumers, and governments. Also, research contributions like the ones of Borenstein et al. [28] or Poon and Sung [29] proof that there are governmental discussions about the legal liability of software developers and designers for violations of the autonomous framework.

Regulatory problems. Despite the lack of regulation for autonomous driving, the problem is a wider one. Actually, despite the lack itself, there are inconsistencies between countries. This will prevent the overarching legislation demanded by experts (cf. expert survey). A general criticism on this is made by Schuelke-Leech et al. [30], Hubbrad [31], and Mackie [32]. Nevertheless, the problem seems to be far more complex. For example, De Bruyne and Werbrouck [33] demonstrate the current incompatibility between EU legislation and autonomous driving. In the example of the EU, alternative solutions are needed to maintain innovative strength.

Thus, previous research has tried to solve the related 'trolley problem' (cf. Section 2.2) to form a moral justifiable basis of decision making. Several approaches exist for this purpose. Islam et al. [34, 35] revealed mathematical models for prioritizing human lives. Awad et al. [7] and Faulhaber et al. [15] attempt to use practical moral evaluations to identify likewise prioritization principles. Furthermore, Aydemir and Dalpiaz [36] defined a five-step method for ethics-aware software engineering that can be also applied to autonomous driving. The steps are based on ethics knowledge, awareness, conscious valuing, and transparency. Also, in practice, there is already progress on this. In Germany, for example a major car manufacturer argued for the protection of the driver rather than the pedestrians. This still lacks on a legal approval [37].

Conclusion and research gap. Despite the agreement on regulatory needs, various problems remain. Especially the different conceptions and missing values on the AD topic seem to be a problem. There is agreement on the need for a consensus solution. For this, a first urgent step is towards the understanding of moral similarities. Appropriate study designs to do so already exist. Nevertheless, they have not yet been transferred into the law-making process.

4.2 Artefact Development

Actual situation. To address this research gap, an understanding of the current law-making process is mandatory. This was one part of our expert survey, yielding the following information: Germany's law-making process for new technologies tries to take on moral similarities and technology assessments by governmental institutes and working groups. These serve as a basis for experts in 'round tables', representing the German society, to prepare new legislative proposals. The legislation itself must be understood as a step-by-step process, confirming the multi-folded approach in [38]. The best consensus from the existing opinions is to be found. This differs significant to the one from the US [39]. Legislation here is subject to greater freedom and is strongly influenced by companies. A legal restriction is made contrary often retrospectively.

Actual problem. Germans' current law-making process for new technologies tends strongly towards the moral similarities of few, rather than many. All experts

https://doi.org/10.30844/wi_2020_a2-wanner

agreed that moral consensus within society should be the real basis. Nevertheless, two of the six experts considered the bias-free collection of such data to be problematic. In particular, a corresponding survey method is not allowed to limit the options to choose from towards one specific kind of ethical approach (e.g., utilitarianism). Moral Machine's study design [7] seems to be valid with our modification by a third option of 'random'. All experts agreed on this, while especially one expert critically remark the strong simplification of the situation in such DS scenarios. The study's results of derived moral similarities should therefore be evaluated critically.

We have developed a corresponding method. The artefact is based on the procedure for the evaluation of hypotheses [40]. It guarantees for a proven development. The knowledge gained from literature and expert interviews has been used to further modify and evaluate our research artefact. Our formed artefact is illustrated in Figure 1.

Overall process. Despite technical assessments, a derivation of moral similarities by a *Moral Consensus Mechanism* must be the first step of Germany's AD law-making process. Thus, it forms the ability to build a consensus. The result will be forwarded to the *Law-making Process*. In this step, it will ease the enactment of the laws forming its (previous missing) basics. The result is a defined *Law-compliant Programming* frame.

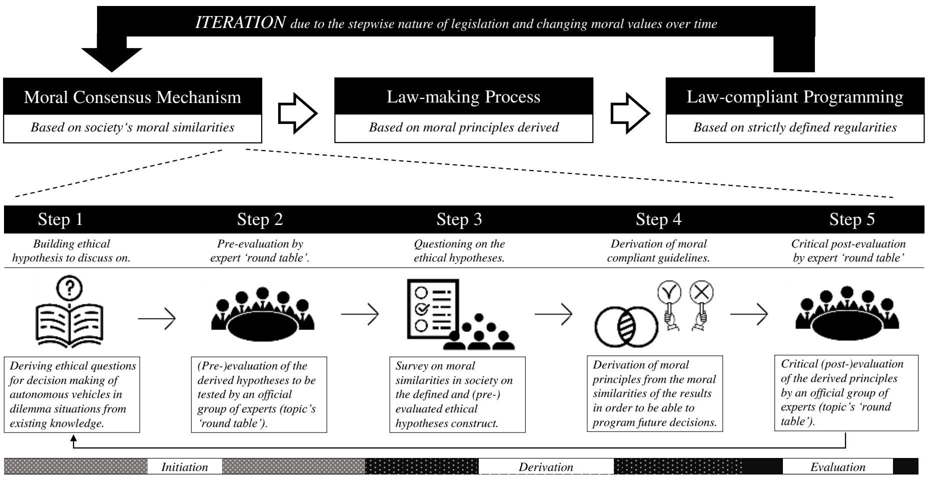

Figure 1. Moral consensus mechanism for German's law-making process in AD

Due to the stepwise development of actual legislation and possible changes in moral convictions, we chose an iterative approach [36].

Step 1: Hypotheses Building. The questioning of moral similarities is always based on a prior selection and thus a limitation of the scope. In consequence, the survey has to build on existing knowledge of ethical principles and moral values. This needs a critical theoretical evaluation resulting in the formulation of hypotheses to test for.

https://doi.org/10.30844/wi_2020_a2-wanner

Step 2: Expert 'Round Table' I. The hypotheses derived and formulated in Step 1 need a (pre-)evaluation by experts. These experts are aware of the moral principles and the technical assessments in parallel. With the overall understanding of the topic they are able to work out the right questions and make variables measurable. The powerful position of the group corresponds with the current law-making procedure in Germany.

Step 3: Society's Questioning. The final scope of hypotheses to test for is translated into appropriate questions depending on the survey's design chosen. It can be either the one of Moral Machine [7] or virtual reality simulations such as [15]. Also, observations of test drives and road traffic can be possible data sources as well. When selecting participants, particular attention should be paid to a consistent cultural background [7].

Step 4: Principle Derivation. In this step, moral principles are derived from the survey's results of the 'correct action' in case of a DS arising. This adapts the successful method of [7]. The moral principles must be universally applicable [17].

Step 5: Expert 'Round Table' II. The derived moral principles are to be finally evaluated by experts. A post-validation with existing survey results or knowledge is recommended. Also, evaluations and comparisons of different cultural backgrounds can be identified in order to form a consensus. If necessary, further iterations follow.

4.3 Evaluation of the Artefact

We evaluated our developed research artefact with several instances: (1) We embodied the governmental working group to develop the preliminary preparations. Further, we designed and conducted the survey. (2) Our interviewed experts formed the expert 'round table'. (3) Society was represented by German participants acquired via social media channels and asked about the DS scenarios defined.

Step 1: Hypotheses Building. First, we tried to critically reviewed the moral principles derived of the study by [7] given only the options of 'left' and 'right' to choose from. The principles were transferred into three hypotheses: (H1) Humans are more likely to be protect than animals. (H2) The number of lives is weighed so that the larger number of people is more likely to be protected. (H3) Young people are more likely to be protected than older people.

Second, based on the study by Sachdeva et al. [41] we stated: (H4) It is more likely to save pedestrians than occupants. Further, Tyler [42] describes the ethical problem of a third instance judging the guilt between a criminal and an innocent person. This is also reflected in traffic. We formulate (H5) to test for: Rule-compliant traffic participants are more likely to be protected than rule-breaking ones.

Step 2: Expert 'Round Table' I. Expect of (H5) all derived hypotheses seem to be of interest. Further, they could be argued to be considered in a process of law-making based on moral similarities. (H5), however, contradicts the general jurisdiction on the one hand. It takes a court to make a judgment about guilt in Germany. On the other hand, especially an implied exchange through facial expressions and gestures seems to be a problem. Other human traffic participant can understand this, but an autonomous vehicle may not. Thus, we exclude this from the

https://doi.org/10.30844/wi_2020_a2-wanner

survey. Further, the planned type of Moral Machine survey and DS situations per hypothesis testing were approved.

Step 3: Society's Questioning. We chose a similar approach to the Moral Machine survey [7]. In addition, the participants had the opportunity to answer with 'random' not to limit our survey design to a specific type of ethical theory (cf. Section 2.3). The hypotheses and measurements from Step 2 were transferred into DS scenarios. Thus, our survey result reveals that 82.3 % of all respondents would (H1) rather kill animals than humans or even themselves. Also, 75.6 % of all respondents decided to (H2) save at least two passengers rather than one driver. Similarly, 72.1 % were more likely to actively kill one person to save at least two more. Our study showed that only 58.0 % of all participants were (H3) more likely to save two children than two elderly people and one child. Only 9.2 % decided against the two children by the factor of the total number of persons. Furthermore, it is (H4) more likely to save pedestrians than occupants. It has been derived out of three different dilemma situations, where almost twice as many survey participants would avoid a pedestrian and kill the vehicle's occupant.

Step 4: Principle Derivation. Based on the data analysis made in Step 3, we can validate the hypotheses (H1) and (H2) of the results of Awad et al. [7]. Thus, the option of 'random' does not seem to lead to any distortion for those. Likewise, the hypothesis (H4) by Sachdeva et al. [41] can be validated as well. For (H3) this differs. It seems quite difficult to decide for. This often leads to the choice of the 'random' option. Although an adjustment of this option shows a clear trend towards younger persons, it must be critically reviewed in the next process step. All four hypotheses form moral derived principles (P) to be evaluated by an expert 'round table'.

Step 5: Expert 'Round Table' II. All experts validated the principle (P1): A human being must be preferred to be protected over animals. Although the result of (P2) is clear by moral similarities, it contradicts German jurisprudence to weight human lives when comparing. This contradicts especially of the German jurisdiction with BVerfG 357/05 paragraph 38. For this reason, the experts consider avoiding an active change of direction. Although the study design made such an answering option possible, it was not chosen. Despite the concerns, the majority of experts were in favour of such a weighing of human lives. Two of them even insist for it in future AD laws. However, a final answer for (P2) needs further research and thus, another iteration cycle. Also, (P3) was discussed controversy. The main problem between the choice for or against young and old is closely linked to (P2). In addition, the values within cultural backgrounds differ. While Western countries tend to value younger people, this is the opposite in Asian ones [7]. As experts recommend finding a world-wide union solution and the moral similarities that were asked for were low, (P3) was rejected. (P4) is already discussed in Faulhaber et al. [15]. They came to the same conclusion, confirming the principle. This is in line with the expert's opinion on that. Nevertheless, the experts note that this is most likely not feasible. If a car is constructed to prefer to save pedestrians before trying to safe occupants, no one will use it. The number of cars sold would fall drastically. Due to the power and importance of car manufacturers in Germany, this would be unthinkable. Thus, we must reject (P4).

https://doi.org/10.30844/wi_2020_a2-wanner

5 Discussion and Limitations

5.1 Methodical Approach

Focusing on the elaborated design of the ethical consensus mechanism, methodological discussion issues, and limitations arise.

Classifying the consensus mechanism into an application field. The iterative approach within the consensus mechanism serves to achieve rapid and continuous integration as well as improvement of autonomous driving within road traffic. Our survey of experts shows that this method is highly suitable. However, all experts interviewed tend to aim for an across-the-board procedure, for example at EU level or in the US, and in the long term even apply it worldwide. In a worldwide development of a framework, different cultural backgrounds must be taken into account, which lead to different principles. Economic factors, infrastructure, and the distinction between individualistic and collectivistic cultures are decisive [7]. Nevertheless, there must be a national idea before a consensus discussion is possible. According to them, in this case the consensus mechanism developed is a suitable approach. Furthermore, the evaluation through our study reveals that the implementation of moral dilemmas in the US is currently handed over to the developer. Despite a loosely stated guideline to comply with ethical principles it is currently assumed that this developer implements moral principles which are not defined. So, even in the US the moral consensus mechanism developed might enable a form of guideline for a sustainable development.

Application of the evaluated uncertainty dilemma. The currently best known and most widespread type of survey for recording moral agreements in the sense of ethical consensus is the Moral Machine questionnaire. An associated problem results from the survey's design. The questions to be answered always require a complete provision of information, which will not always be given in practice. It can therefore be assumed that the rules are not always fully applicable. This is also based on the differentiation between the 'right decision' making at real-life interaction and hypothetical interaction at questionnaires. So the Institute of Cognitive Science at the University of Osnabrück reveal that subjects with and without time pressure make asymmetric decisions [43]. The post-evaluation of derived principles should take into account that people react differently in hypothetical scenarios than in real scenarios. People in real scenarios react more emotionally and think more rationally in abstract scenarios [44]. As our consensus-mechanism demonstrate and the interview study confirms, the principles from the dilemmas are nevertheless required as a first step within an iterative approach. Thus, for instance, a roundtable of experts within the German jurisdiction prepare reports, which serve as a basis for legislation. So, these experts represent the views of the public and thus of the surveyed individuals. On the other hand, within the USA, through industry, it is a very innovation-driven approach in the law-making process. As a result, the interests of the customers and thus of the public are indirectly integrated in the decision-making process in terms of the willingness of customers to buy from car manufacturers.

https://doi.org/10.30844/wi_2020_a2-wanner

5.2 Ethical Derivations

In view of the elaborated design of the ethical consensus mechanism, corresponding ethical discussion issues and limits arise.

Ethical foundation. Consequentialists call for damage-minimizing programming of autonomous vehicles and argue that 90 % of traffic accidents are due to human error, which is thus avoided [45]. However, such a system does not consider the rights of the individual. This violates ethical codes like the one of the Institute of Electrical and Electronics Engineers (IEEE) to treat all persons fairly and not to discriminate related to personal characteristics such as religion, gender or age. The deontological approach corresponds to those of current jurisprudence in states like Germany. Here, maximizing intersubjective benefits does not outweigh the violation of fundamental rights. For example, Germany's BVerfG 357/05 about the non-legitimation to fire on a flying object, which is used against human life due to guarantee of human dignity stated in §1(1) in Germany's constitutional law. This is also manifested in the 'trolley problem'. For consequentialists, which follow an utilitaristic approach, a weighting between human lives is an approach to generating 'greatest happiness of the greatest number'. Deontologists, due to the rights of the individual, criticize this decision [46]. By adding the option of 'random' for our Moral Machine survey study, we avoided this preliminary and allowed the participants to follow any of those ethical principles.

Post-evaluation of (P4). The preferred protection of pedestrians against occupants (P4), contradicts Daimler's statement to protect the occupants first [37]. The identity of those involved in a moral dilemma situation is not clear at the time of programming. In consequence, it is possible that all individuals can be pedestrians as well as occupants. Those who, due to their need for mobility, introduce risks into society in connection with vehicles also must bear the responsibility for this. Mercedes-Benz executive Christoph von Hugo [37] explains that it makes more sense to protect the occupants, since they are known as opposed to pedestrians. However, current findings seem to indicate that this is the exception at Daimler. On the other hand, our interview study reveals that the acceptance for the customer of an autonomous vehicle would be increased by this circumstance, when he knows that there is no obligatory willingness to sacrifice his own life to save others.

Misconduct. Even when autonomous vehicles drive perfectly, accidents cannot be ruled out. The transition to purely autonomous vehicles on the road is very likely to become fluid. The autonomous vehicle must adapt to mixed traffic [47]. Wild animals and pedestrians continue to pose sufficient danger potential in road traffic. A consideration between occupants and pedestrians is not necessary from this point of view, as the question of misconduct is first considered (P5).

Creation of Misconduct. An AD that applies moral principles can be leveraged by third parties for their own benefit through the targeted use of these principles. Goodall describes scenarios in which pedestrians, knowing that they are spared due to their group size (P2 and P4), carelessly cross a busy road [48]. One possible approach to counteracting this is the integration of rule-compliant behavior (P5) [47].

https://doi.org/10.30844/wi_2020_a2-wanner

6 Conclusion and Outlook

The PoC procedure applied according to [8] had the first step to gain an understanding of the problems involved in the development of autonomous vehicles in connection with jurisdiction and consequently, to present a possible approach for implementation. The goal of this paper, based on [8], is to provide a practical and scientific basis for a proof-of-use. The findings in the PoC are summarized in the following.

The review of the existing knowledge of research in the section of ethical AD (cf. Section 3.2) revealed that although it would already be technically possible to use self-driving vehicles in today's traffic, practical application would fail due to the lack of a legal framework to program the AI-logic component. Despite initial efforts in the direction of legislation, there is a lack of appropriate foundations for the development of adequate laws. Our research artefact addresses the problem and provides guidance to solve it by developing a consensus mechanism to derive concrete principles from moral similarities regarding decisions in DS as a fundamental basis for law-making and AI-programming.

An exemplary implementation of the developed five-step approach to consensus finding was carried out in a quantitative study modelled on behalf of the design of the Moral Machine experiment. This confirms the principles derived earlier by a research group [7]. In addition, two predefined new hypotheses could be developed and one of them validated as derived principle for 'right action' of an AD vehicle in a specific dilemma situation. Future research efforts could start here and test the new principle derived, or test the validity with regard to other cultural backgrounds as participants differ in moral concepts [7]. Furthermore, other problem situations can also be questioned to develop new findings and derive new principles. In addition, critical scrutinizes of most methodologically appropriate design of the step concept must be carried out for each part, with corresponding recommendations. Overall, however, our approach proves that the consensus mechanism presented can provide an essential guide for further research.

References

1. Bain & Company, https://www.bain.com/de/ueber-uns/presse/pressemitteilungen/germany/2019/bain-analyse-zum-mobilitatssektor/ (2019)
2. Rosenzweig, J., Bartl, M.: A review and analysis of literature on autonomous driving. The Making of Innovation. E-Journal (2015)
3. Merfeld, K., Wilhelms, M.-P., Henkel, S., Kreutzer, K.: Carsharing with shared autonomous vehicles: Uncovering drivers, barriers and future developments. Technological Forecasting and Social Change 144, 66-81 (2019)
4. Marshall, A., https://www.wired.com/story/uber-self-driving-car-crash-arizona-pedestrian/ (2018)

https://doi.org/10.30844/wi_2020_a2-wanner

5. Deutscher Bundestag: Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes. Fundstelle: Beschlussempfehlung und Bericht des Ausschusses für Verkehr und digitale Infrastruktur. (2017)
6. Janiesch, C., Fischer, M., Winkelmann, A., Nentwich, V.: Specifying autonomy in the Internet of Things: the autonomy model and notation. Information Systems and e-Business Management 17, 159-194 (2019)
7. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I.: The moral machine experiment. Nature 563, 59 (2018)
8. Nunamaker Jr, J., Briggs, R., Derrick, D., Schwabe, G.: The last research mile: Achieving both rigor and relevance in information systems research. Journal of management information systems 32, 10-47 (2015)
9. Moor, J.: The nature, importance, and difficulty of machine ethics. IEEE intelligent systems 21, 18-21 (2006)
10. Maurer, M., Gerdes, J., Lenz, B., Winner, H.: Autonomes Fahren: technische, rechtliche und gesellschaftliche Aspekte. Springer-Verlag, Berlin (2015)
11. Verband der Automobilindustrie e.V.: Von Fahrerassistenzsystemen zum automatisierten Fahren - VDA. Brandenburgische Universitätsdruckerei. (2015)
12. Litman, T.: Autonomous vehicle implementation predictions. Victoria Transport Policy Institute, Victoria, Canada (2017)
13. Bellet, T., Cunneen, M., Mullins, M., Murphy, F., Pütz, F., Spickermann, F., Braendle, C., Baumann, M.F.: From semi to fully autonomous vehicles: New emerging risks and ethico-legal challenges for human-machine interactions. Transportation research part F: traffic psychology and behaviour 63, 153-164 (2019)
14. Nyholm, S., Smids, J.: The ethics of accident-algorithms for self-driving cars: An applied trolley problem? Ethical theory and moral practice 19, 1275-1289 (2016)
15. Faulhaber, A., Dittmer, A., Blind, F., Wächter, M., Timm, S., Sütfeld, L., Stephan, A., Pipa, G., König, P.: Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles. Science and engineering ethics 25, 399-418 (2019)
16. Kezar, A.: Leadership for a better world: Understanding the social change model of leadership development. The Journal of Higher Education 81, 670-671 (2010)
17. Noll, B.: Wirtschafts-und Unternehmensethik in der Marktwirtschaft. Kohlhammer Verlag, Stuttgart (2013)
18. Gerdes, J., Thornton, S.: Implementable ethics for autonomous vehicles. Autonomes fahren, 87-102. Springer, Berlin (2015)
19. Mill, J.: Utilitarianism and on liberty: Including Mill's' Essay on Bentham'and selections from the writings of Jeremy Bentham and John Austin. John Wiley & Sons, New Jersey, UK (2008)
20. Pleger, W.: Das gute Leben: eine Einführung in die Ethik. Springer-Verlag, Heidelberg (2017)
21. Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A design science research methodology for information systems research. Journal of management information systems 24, 45-77 (2007)

https://doi.org/10.30844/wi_2020_a2-wanner

22. Webster, J., Watson, R.: Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly xiii-xxiii (2002)
23. Myers, M., Newman, M.: The qualitative interview in IS research: Examining the craft. Information and organization 17, 2-26 (2007)
24. Schnell, R., Hill, P., Esser, E.: Methoden der empirischen Sozialforschung. Oldenburger Wissenschaftsverlag, Oldenburg (1999)
25. Cunneen, M., Mullins, M., Murphy, F.: Autonomous vehicles and embedded artificial intelligence: The challenges of framing machine driving decisions. Applied Artificial Intelligence 33, 706-731 (2019)
26. Baumann, M., Brändle, C., Coenen, C., Zimmer-Merkle, S.: Taking responsibility: a responsible research and innovation (RRI) perspective on insurance issues of semi-autonomous driving. Transportation Research Part A: Policy and Practice 124, 557-572 (2019)
27. Li, G., Deng, X., Gao, Z., Chen, F.: Analysis on Ethical Problems of Artificial Intelligence Technology. Proceedings of the 2019 International Conference on Modern Educational Technology, 101-105. ACM (2019)
28. Borenstein, J., Herkert, J., Miller, K.: Self-driving cars: Ethical responsibilities of design engineers. IEEE Technology and Society Magazine 36, 67-75 (2017)
29. Poon, N., Sung, J.: Self-driving cars and AI-assisted endoscopy: Who should take the responsibility when things go wrong? Journal of gastroenterology and hepatology 34, 625-626 (2019)
30. Schuelke-Leech, B.-A., Jordan, S., Barry, B.: Regulating Autonomy: An Assessment of Policy Language for Highly Automated Vehicles. Review of Policy Research (2019)
31. Hubbard, S.: Automated vehicle legislative issues. Transportation research record 2672, 1-13 (2018)
32. Mackie, T.: Proving liability for highly and fully automated vehicle accidents in Australia. Computer Law & Security Review 34, 1314-1332 (2018)
33. De Bruyne, J., Werbrouck, J.: Merging self-driving cars with the law. Computer law & security review 34, 1150-1153 (2018)
34. Islam, M., Rashid, S.: Algorithm for Ethical Decision Making at Times of Accidents for Autonomous Vehicles. 4th International Conference on Electrical Engineering and Information & Communication Technology, 438-442. IEEE (2018)
35. Peterson, M.: The value alignment problem: a geometric approach. Ethics and Information Technology 21, 19-28 (2019)
36. Aydemir, F., Dalpiaz, F.: Poster: Ethics-Aware Software Engineering. 40th International Conference on Software Engineering, 228-229. IEEE (2018)
37. Morris, D., http://fortune.com/2016/10/15/mercedes-self-driving-car-ethics. (2016)
38. Winfield, A., Jirotka, M.: Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376, 1-13 (2018)

39. Shladover, S., Nowakowski, C.: Regulatory challenges for road vehicle automation: Lessons from the california experience. Transportation Research Part A: Policy and Practice (2019)
40. Campbell, D., Stanley, J.: Experimental and quasi-experimental designs for research. Ravenio Books, London (2015)
41. Sachdeva, S., Iliev, R., Ekhtiari, H., Dehghani, M.: The role of self-sacrifice in moral dilemmas. PloS one 10, 1-12 (2015)
42. Tyler, T.: Restorative justice and procedural justice: Dealing with rule breaking. Journal of social issues 62, 307-326 (2006)
43. Sütfeld, L., Gast, R., König, P., Pipa, G.: 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, 122 (2017)
44. Bostyn, D., Sevenhant, S., Roets, A.: Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. Psychological science 29, 1084-1093 (2018)
45. Bonnefon, J.-F., Shariff, A., Rahwan, I.: The social dilemma of autonomous vehicles. Science 352, 1573-1576 (2016)
46. Hevelke, A., Nida-Rümelin, J.: Selbstfahrende Autos und Trolley-Probleme: Zum Aufrechnen von Menschenleben im Falle unausweichlicher Unfälle. Jahrbuch für Wissenschaft und Ethik 19, 5-24 (2015)
47. Matthaei, R., Reschka, A., Rieken, J., Dierkes, F., Ulbrich, S., Winkle, T., Maurer, M.: Autonomes Fahren. Handbuch Fahrerassistenzsysteme, 1139-1165. Springer, Wiesbaden (2015)
48. Goodall, N.: Ethical decision making during automated vehicle crashes. Transportation Research Record 2424, 58-65 (2014)

https://doi.org/10.30844/wi_2020_a2-wanner