

Distributive Justice as an Ethical Principle for Autonomous Vehicle Behavior Beyond Hazard Scenarios

Manuel Dietrich, Thomas Weisswange

2019

Preprint:

This is an accepted article published in Ethics and Information Technology. The final authenticated version is available online at:
<https://doi.org/10.1007/s10676-019-09504-3>

Distributive Justice as an Ethical Principle for Autonomous Vehicle Behavior Beyond Hazard Scenarios

Manuel Dietrich · Thomas H. Weisswange

Received: 13 June 2018 / Accepted: 3 May 2019

Abstract Through modern driver assistant systems, algorithmic decisions already have a significant impact on the behavior of vehicles in everyday traffic. This will become even more prominent in the near future considering the development of autonomous driving functionality. The need to consider ethical principles in the design of such systems is generally acknowledged. However, scope, principles and strategies for their implementations are not yet clear. Most of the current discussions concentrate on situations of unavoidable crashes in which the life of human beings is existentially affected. In this paper, we argue that ethical considerations should be mandatory for any algorithmic decision of autonomous vehicles, instead of a limitation to hazard situations. Such an ethically aligned behavior is relevant because autonomous vehicles, like any other traffic participants, operate in a shared public space, where every behavioral decision impacts the operational possibilities of others. These possibilities concern the fulfillment of a road-user's safety, utility and comfort needs. We propose that, to operate ethically in such space, an autonomous vehicle will have to take its behavior decisions according to a just distribution of operational possibilities among all traffic participants. Using an application on a partially-autonomous prototype vehicle, we describe how to apply and implement concepts of distributive justice to the driving environment and demonstrate the impact on its behavior in comparison to an advanced but egoistic decision maker.

Keywords Autonomous Vehicles · Distributive Justice · Ethics · Responsibility · Self-driving Cars · Non-Hazard

1 Introduction

Ethical concepts for the design of autonomous decision-making systems, in particular in the area of self-driving cars, have become a flourishing research topic. Having fully autonomous vehicles on public roads is considered a realistic scenario in the nearby future, as indicated by intense work in this field, including real-world fleet testing done in the US, China and recent plans for Europe.¹

Researchers from different disciplines, including computer science, law, philosophy, and neuroscience have published their concepts, ideas or concerns on implementing ethics for autonomous decision systems (see Sec. 2 for references). Indeed, it is to be expected, that the challenges can not be tackled sufficiently from a single point of expertise. It is mandatory to mutually exercise a substantial ethical discussion and at the same time consider the existing and future technical capabilities in sensing and traffic scene interpretation. We contribute towards this by providing an ethical discussion on distributive justice for the road environment together with an implementation of the concepts within a semi-autonomous highway driving application.

In our approach, we argue for the relevance of an ethically aligned behavior of autonomous vehicles beyond hazard scenarios. In recent years, many authors

Manuel Dietrich
Honda Research Institute Europe GmbH,
Offenbach, Germany
E-mail: manuel.dietrich@honda-ri.de

Thomas H. Weisswange
Honda Research Institute Europe GmbH

¹ See for example Waymo's self-driving cars (Waymo (2017)) or Volvo's *Drive Me* plans for alpha-testing with real families (<https://www.volvocars.com/intl/about/our-innovation-brands/intellisafe/autonomous-driving/drive-me>).

have done important work to show that ethical considerations are relevant for designing autonomous vehicles (Goodall (2014a), Goodall (2014b), Lin (2015)). Most existing work focuses on situations in which a serious, life-threatening crash is unavoidable. The relevance for extending ethics to general non-hazardous driving can be justified by the fact that life-threatening situations do not exist in isolation. In fact most driving behavior decisions are taken within a shared operational space which means that traffic participants act to serve their utility, comfort and safety needs but at the same time their behavior reshapes the operational space of others – which includes an impact on the others’ chance to get into a hazardous situation.

We discuss in the paper how autonomous vehicles should be designed to respond to the ethical significance of decisions they will have to make in traffic. Furthermore, we think the decision rules which align the behavior of autonomous cars should be specified with respect to concepts of distributive justice. Applying distributive justice as behavior principle is presumed to lead to decisions which result in a fair allocation of driving opportunities between all traffic participants.

In the paper, we first discuss concepts of justice, originally targeting the institutional distribution of social goods (see Sec. 4). Subsequently, we map the general principle of distributed justice onto an ethical formulation for algorithmic decision-making in traffic situations (see Sec. 5). We also consider in more depth, why applied distributive justice should be taken care of in designing general autonomous driving systems. The effects of the concept on real world decision systems will be illustrated through an implementation in a prototypical application for autonomous longitudinal and lateral autonomous highway driving (see Sec. 6). It is investigated how the vehicle would decide based on the justice principle as compared to a more common concept that prioritizes the vehicles and therefore the passengers self-interest (see. Sec. 7). We will end the paper with a discussion of the results and a conclusion (see Sec. 8).

2 Related Work

Within the last years researchers from different disciplines have approached the ethical challenges of future self-driving scenarios. With self-driving cars, advanced machine learning systems which are commonly named as artificial intelligence technologies will visibly enter the public space where every citizen will be affected. Besides the discussion of ethical principles for implementation, questions concerning the responsibility of drivers and manufacturers are discussed (Hevelke

and Nida-Rümelin (2015), Coeckelbergh (2016), Charisi et al (2017)). The literature on responsibility will be relevant in Sec. 5.

Most researchers have approached the ethics of autonomous vehicles with respect to situations of *unavoidable* crashes. Goodall says that the goal for implementing ethics into autonomous vehicles is to find the “best way to crash” (Goodall (2014a), p. 60). Similarly Nyholm and Smids speak about “accident-algorithms” for situations in which “collisions are highly likely or unavoidable” (Nyholm and Smids (2016), p. 1275). Goodall argues that choosing a path in a crash situations is an “activity that transfers risk from one person to another” (with the potential outcome that humans are seriously harmed) and therefore is something which should involve ethical considerations (Goodall (2014b), p. 5).

To apply ethical concepts for autonomous vehicles to enable decision-making in crash situations, the principles have to be formalized. For that, many authors use textbook examples to illustrate how ethics can be implemented within an algorithm. An approach by Gerdes and Thornton (Gerdes and Thornton (2016)) states an analogy between the moral decisions made following the ethical principle of utilitarianism and the concepts of utility (inverse: cost) optimization (“cost optimal control”). This is straightforward because the utilitarian principle *is* a way to frame ethics as utility optimization: it targets to choose the option with highest average utility between all relevant persons. Additionally, they translate deontological ethics into what is called constraints in the cost optimal control terminology which means that the cost of a certain predicted outcome is determined as infinite, so that the options not having a reasonable chance to produce this outcome are always favored. If for all options human lives are in danger, as it is the case for the classic trolley-problem², deontology ethics which values human life as the highest worth (infinite costs), can not help to discriminate between different life-threatening options. The approach of using simplified ethical concepts, as Gerdes and Thornton do, generally lacks a deeper discussion on the applications and parameters and clear reasons why they should be suitable for the driving setting or certain driving situations.

² The trolley-problem is a classic ethical thought experiment, which is referred to in the ethics of autonomous vehicles. It can be described as follows: A driverless trolley is heading towards a group of people on a track. The only chance to save their lives is to pull a switch which has the effect that the trolley is redirected to a side-track where only a single person is killed instead. This scenario is used to compare ethical justifications in dilemma situation.

Another way to approach the ethics of crashing is by discussing ethical theories of social justice in the context of autonomous vehicles. For example Leben refers to Rawls' famous fairness approach (Rawls (2009)). Leben's motivation is that the previously-mentioned utilitarian approach is not convincing for the ethics of crashing, especially "for dealing with trolley-style dilemmas" (Leben (2017), p. 107). He focuses on Rawls' second-order principle of just distribution, namely "the difference principle". The difference principle targets a distribution of goods so that the socially disadvantaged people with respect to certain criteria of social rank, wealth or general living conditions, profit the most (favors the welfare of the worst-off people in a society). Leben applies the principle to the *distribution* of harm in life-threatening crash situations. Harm is rated by a "single scale of severe injury, with death being the most extreme point on this scale" (Leben (2017), p. 113). The idea is that in a serious crash situation, the option is preferred for which is true that the most extreme position of a single person on the injury-scale is minimal.

A social justice approach which goes beyond driving decisions in unavoidable accidents has been discussed by Mladenovic und McPherson, who argue on the level of road traffic control with self-driving cars, rather than on the individual autonomous vehicle behavior (Mladenovic and McPherson (2016)). Traffic flow control is not the scope of our approach but the arguments why social justice is of relevance for the road environment is useful for our approach. Mladenovic und McPherson especially focus on a general worth or even human right of free movement, so that the individual needs and interests related to the freedom of movement on a public road have to be distributed fairly.³ As needs and interest of traffic participants (referred to as desiderata), they specify "safety, sustainability, privacy, efficiency, and equality of access" (Mladenovic and McPherson (2016), p. 1135). They argue for a hierarchically distributed self-organizing system. That means they imagine an interchange between the micro-level (decentralized vehicle control) and a macro-level (traffic control): "background system of rules for managing such cooperation, in order to maximize the benefits from self-organization and prevent system failure" (Mladenovic and McPherson (2016), p. 1142). One practical downside of their approach is that such a system would require a fully connected driving infrastructure with every traffic participant taking part in it.

³ "Protecting the human right to free movement while ensuring that the opportunities to meet human needs and interests are fairly distributed" (Mladenovic and McPherson (2016), p. 1135).

Our approach contains elements of both of the two previous approaches, by suggesting a decentralized control of vehicles based on distributive justice as it is approached by Leben but with focus on non-hazardous behavior decisions as it is approached by Mladenovic und McPherson.

3 Discriminating Hazard from Non-Hazard Situations

We characterize hazardous traffic situations as circumstances where humans involved, e.g in their cars or as pedestrians, have an immediate and significant chance to be harmed. In extreme cases hazard situations are life-threatening as it is primary discussed in the ethics of crashing literature (see Sec. 2). All driving situations in which the health of humans is not affected (with a significantly probability), we consider as non-hazard situation.

An ethical consideration of hazardous situations will not be part of a deeper analysis because it is not the focus of the paper. However, most people will agree that human life has the highest value and cannot be accounted with other values for instance property damage or animal life. The democratic rights of individuals go even further, so that there is no trade-off between humans allowed: "the deontological character of a liberal democracies normative order manifests itself partly in preventing trade-offs when it comes to certain individual rights and liberties" (Hevelke and Nida-Rümelin (2015), p. 655). This principle of humanity forbids to treat people differently with respect to their characteristics (e.g. age, sex, bodily and mental state).⁴ In most countries of the world the principles of humanity are part of the constitution.

Many scientific approaches discussing the ethics of self-driving cars are based on experiments on how humans would decide in *unavoidable* crash situations (often constructed as trolley-cases. In the experiments the participants are usually requested to decide for a crash option whereby the options differ regarding the quantity and the personal characteristics of the people who will be killed (so called *experimental ethics* (Lütge et al (2014)) are for instance done by Bonnefon et al (2016) or Nyholm (2018b). If the results of such experiments would be implemented in autonomous vehicles – the preferences regarding who ought be killed and who not –, it would be a decision rule which treats people differently with respect to their characteristics and so violate

⁴ For example the results of the German "Ethics Commission" strongly stresses this fundamental right with respect to autonomous driving (Ethics-Commission (2017)). See Luetge (2017) for comments on the report.

fundamental human rights. A specific critique on these approaches is already formulated by Keeling (2017) as well as Nyholm and Smids (2016).

We think these are good reasons to argue for a decision scheme which minimizes the possibility of fatality without a qualitative discrimination between humans at risk. We have to consider non-hazardous situations differently, which we will discuss in the next section. Tab. 1 summarizes this basic distinction.

4 Basics: Four Dimensions of Distributive Justice

Most theories of social justice can be described by a variation of the following four key dimensions of just distribution.⁵

1. Distributive goods/burdens: e.g. wealth, education, healthcare; rights and liberties (goods); military service (burdens)
2. Recipients of the distribution: e.g. members in society, future generations, global perspective, *local justice*
3. Characteristics of the recipient which are considered as relevant for the distribution: e.g. merit, need, ability, guilt
4. Basis of distribution (rule): e.g. strict equality, maximizing utility, difference principle

Distributive justice covers the allocation of social goods as well as burdens that people (citizens) have in society.⁶ The just distribution is relevant for goods which are seen as central values for people in society. Rawls calls the goods which are qualified for just distribution “primary goods”⁷, others, for example Finnis, talk about “common good[s]”.⁸ Distributive justice is mainly discussed as an ethical decision-base for social institutions (e.g. universities, healthcare institutions)

⁵ Lamont and Favor (2017) presents a similar set of key aspects (see chapter: “Scope and Role of Distributive Principles”).

⁶ One definition: “justice in distribution of good and evil” in a “comparative treatment of individuals” (Frankena (1963), p. 10).

⁷ “Things that every rational man is presumed to want”, for example: “liberty and opportunity, income and wealth, and the social bases of self-respect” (Rawls (2009), p. 54).

⁸ He defines common goods as “resources, opportunities, profits and advantages, roles and offices, responsibilities, taxes and burdens - in general, the common stock and the incidents of communal enterprise, which do not serve the common good unless and until they are appropriated to particular individuals” (Finnis (2011), p.166).

with respect to the distribution of material and non-material goods (e.g. education) made under a “scarcity of resources” (Schmidt (1992), p. 8).

A second key attribute is the scope of the justice theory, that means for which domains the concepts are mandatory. General approaches, for example Rawls’ theory, aim for an “overall design of the basic structure of society” (Schmidt (1992), p. 2). These compete with local justice theories, which highlight a high domain dependency on what is a just distribution (Elster (1993), Schmidt (1992)).

Concepts of distributive justice also differ in what characteristics of people should be taken into account (e.g. the members of society as recipients of the distribution). A common goal for a just distribution is to compensate involuntary disadvantages of citizens. For instance Cohen takes this position and states the elimination of “involuntary disadvantage, [...] disadvantage for which the suffer cannot be held responsible” as main target for social justice (Cohen (1989), p. 916). Such disadvantages are for instance unfavorable natural disposition, misfortune or bad luck.⁹ On the other hand, unfortunate life situation which are seen as self-caused should not be taken into account for a compensatory distribution of social goods. Often it is also argued that social merits or efforts (above average) should be worth it for the individual and be rewarded with a higher share of certain goods (Dworkin (1981), Cohen (1989)). Another stream of social justice theories dismiss any fault- or merit-dependent distribution below a certain humanistic threshold (absolute standards). That means certain goods for instance basic health-care, essential food and shelter should be guaranteed unconditionally (Nussbaum (1992), Margalit (1998)).

One central topic of controversy is the legitimate *basis of distribution*. There exist three dominant principles, which we will focus on, although there are many more including combinations of those three.

Strict equality theorists argue for a distribution of goods (or burdens) which leads to an arithmetically balanced share of goods for all recipients.¹⁰ For instance the vision of a society-wide basic income meets this principle: a fixed amount of money is granted to every citizen independent of any factor (needs or merits) which distinguishes persons (citizens) from each other. Another example for a strict equal distribution in the field of transportation is the deployment of a free public transport system.

⁹ This perspective has become popular as “luck egalitarianism” (Anderson (1999), p. 327).

¹⁰ For instance Persson argues for an extreme egalitarianism position (Persson (2007)).

	Hazard Situations	Non-Hazard Situations
Definition	All driving situations where humans have an immediate and significant chance to be harmed	All driving situations where humans have no immediate and significant chance to be harmed
Goods	Human Life	The public road, i.e. the objectives and conditions of usage e.g. comfort, safety, utility
Recipients of distribution	Anyone in situation of immediate risk	Any traffic participant (in-situ)
Characteristics of recipients	No differentiation in treatment allowed	Equality of treatment - exceptions could be possible (E.g. weighting needs of an ambulances higher)
Distribution Rule	Minimizing fatalities	Distributive justice (Fairness)

Table 1 In this table we compare hazard situations and non-hazardous situations. For the non-hazard type (right side) the relevant criteria with respect to the dimensions of social justice are summarized. On the left side, the same dimensions are used to arrange the primary results which were argued for with respect to the discussion of hazardous situations in Sec. 3

Applying the **utilitarian** principle leads to distributive decisions which result in an in sum highest benefit for all recipients of the distribution (Mill (1861), Bentham (1996), Harsanyi (1977)). Implicit in the application of this principle is that individuals or certain groups in society can be put at a disadvantage as long as the benefits for the majority increase.¹¹ Infrastructural investments which target towards the highest in average fulfillment of transport demands would lead to a well developed metropolitan transport and omit rural areas.

Applying the **difference principle** leads to an allocation of goods in which the most disadvantaged recipients, introduced by Rawls as the “worst-off” citizens in society, are supported the most (Rawls (2009)). The approach is also termed *justice as fairness* because, according to Rawls, the principle provides a base for enduring cooperation among citizens. In Rawls’ theory the difference principle is only the second subordinated principle for a just society. Equality as equal liberty and equality of opportunity is the first one. With respect to the transportation example, one could argue, based on this principle, that people who are less privileged (e.g. low income) should ride free of charge to allow them to participate more in social life.

In the next section, we will focus on the difference principle and show how it can be applied to autonomous driving. The other two presented principles will be also briefly discussed with respect to their applicability.

5 Distributive Justice for Autonomous Driving Beyond Hazard Situations

In a driving situation a human driver or an assistant system has to decide between different options, for example braking, acceleration or lane-change maneuvers. Options are weighed based on in-situ perception respecting various driving factors, whereby one of the factors is the likelihood of an accident. Other relevant *forces* are utility factors, for instance the progress towards a certain destination or driving comfort. In most circumstances an individual’s driving needs are in conflict with those of other road participants due to the limited road space.

We argue in the paper that most decisions during public driving are of “normative significance” (Misselhorn (2015), p. 19). By entering the road, people enact their right of free movement and at the same time join a community in which the participants are part of a shared operational space. Participating in traffic is an interpersonal and collaborative activity. Persons who act freely (expressing freedom) in such a space take personal risks and at the same time shape the shared risk-space.¹² The normative significance of behaviors which limit or enable others’ operational opportunities is not only justified by the fact that others could be brought in to situations with higher or lower risks (which influences the chance of being involved in an accident). Though, every behavior decision has an impact on others’ opportunities to make use of the road to fulfill their comfort and utility needs.

Distributive justice becomes relevant when there is a social good/burden which has to be allocated between citizens (previous section). For the driving scenario, the

¹¹ “Imagine a case in which the enslavement of a very small minority population will make the much larger majority population very happy” (Shafer-Landau (2007), p. 415).

¹² For instance Ferretti argues that any “risk imposition” is of ethical relevance even when in the end no actual damage occurs (Ferretti (2016)).

road can be seen as such shared social good – the interrelated and conflicting objectives and conditions of usage have to be mediated by distribution principles. As mentioned in the related work section, Leben argues in a similar direction, but focuses on unavoidable crash cases. He says that autonomous cars will face decisions which are “morally relevant” because in case of an accident their behavior has major impact on the “distribution of health and survival” (Leben (2017), p. 3). He thinks that distributive justice as fairness is an appropriate framework to align such decisions. In our approach we address a different sphere of moral values which are also relevant in the context of driving. Not the immediate impact on human health but rather the expected likeliness to get into a dangerous situation (risks) together with conflicting possibilities of using the road are considered for distribution.¹³

If most driving decisions on public roads are normatively significant, what are the consequences with respect to the design or regulation of autonomous vehicles? Human drivers make moral decisions in non-hazard traffic situations: For example they give way by accepting the extra effort of changing the lane or insist on their right of way when they are in a hurry. From an ethical standpoint, it is their moral right to act freely to a certain degree, that means they are allowed to act in a certain self-interested way. On the other hand, they can be held responsible for their actions (for instance when self-interested behavior is considered as careless).

Most authors argue that autonomous vehicles are as well capable to behave agent-like but the type of agency differs (compare: Coeckelbergh (2016), p. 752, p. 756, Nyholm (2018a)). Nyholm introduces a useful distinction between different types of agency which he already relates to the capabilities of autonomous vehicles. Following his approach, the most advanced form of agency, autonomous vehicles are able to perform, is “domain-specific principled agency” (Nyholm (2018a), p. 1208). Being able to act this way means to pursue goals on basis of representations which are regulated by certain rules, within certain limited domains (Nyholm (2018a), p. 1208).

This type of agency fits to common approaches of autonomous driving systems: These systems are able to represent a driving scene with its dynamics via environmental sensing and are able to make predictions about the scene and align their behavior accordingly. A rule-set, for instance including estimations of the value or disvalue of certain behavioral consequences (severity of an accident) (Jenkins (2017)), allows to decide for a

¹³ Leben is also aware of that and claims a more general applicability of his approach in the conclusion of his article (Leben (2017)).

best option.¹⁴ We will discuss this in more detail from a practical point of view in Sec. 6.

A more advanced form of agency, which has so far been reserved for human actors is “responsible agency” (Nyholm (2018a), p. 1208). It contains the ability of the actor to understand criticism along with the ability to defend one’s actions. Additionally, these agents are socially embedded and can be punished which means that they are candidates for “retributive blame” (Danaher (2016), p. 2). As autonomous vehicles decide autonomously in their domain (of traffic) they are causally responsible for behavioral outcomes but not morally responsible (a distinction different authors come up with: Danaher (2016), Misselhorn (2015)).

As we said before, we do understand autonomous vehicles as domain-specific principled agents that means decisions are aligned by explicit principles. Furthermore, we argued that most decisions they have to make in traffic are of normative significance. As result of bringing both together, we can say that autonomous vehicles which are able to perform *principled agency* should be created by taking ethical considerations into account for defining these principles. Even more, so far the specification of such principles lay within the designers choice, it is the designers’ obligation to care for ethics. It is part of the “forward-looking responsibility” product designers have (Nyholm and Smids (2016), p. 1283).¹⁵

Assuming that programmers are responsible for behavior principles does not mean that they account for everything which might happen when the technology is in use. Even when one can say that autonomous systems can decide consistently based on the data available, sensing limitations and unexpected traffic situations, which are very different from common expectations, can make the behavior in the real-world partly unpredictable.¹⁶

¹⁴ In some autonomous vehicle prototypes, behavioral decisions are made by a “black-box” machine learning algorithm, that maps incoming information (e.g. image data) to behavior output based on prior training examples (Bojarski et al (2016)).

¹⁵ In a report of the ethic-commission, assigned by the German Ministry of Transport, the authors claim that autonomous vehicles should be programmed that they “drive in a defensive and anticipatory manner, posing as little risk as possible to vulnerable road users” (Ethics-Commission (2017), p. 6).

¹⁶ Researchers have pointed at a potential responsibility gap with respect to autonomous systems (Danaher (2016), Sparrow (2007)).

5.1 Ways to Apply Distributive Justice to Align Autonomous Decisions in Traffic

Since we have argued that the road is a social good which falls under the obligation of just allocation, from an autonomous vehicle's decision perspective the interests (driving needs) of the owner/driver are no longer the dominating parameters. Consequently, we think that autonomous vehicles should make driving decisions by taking all the other traffic participants into account. We argued that the share of road, which individuals can gain, is conditioned upon their driving opportunities. According to that, autonomous vehicles should take others into account by making behavior decisions so that the driving opportunities of the traffic participants in a scene are justly distributed.

In order to move towards applying a justice principle for autonomous driving decisions, we have to find ways to quantify driving opportunities. As mentioned before, three major factors, namely safety, utility and comfort are considered. In the next section (Sec. 6, esp. Tab. 2), we discuss in more detail which features can be used to estimate this factors.

We think the **difference principle** is an appropriate maxim, when it comes to the distribution of driving opportunities. At the end of the section, we will also briefly discuss alternative distribution rules with respect to their applicability to driving situations beyond hazard.

With respect to Rawls, who introduced the difference principle, for a just society it is most important that citizens have equal liberties and rights. One basic right of citizen is that they are treated equally in accordance with the rule of law (Rawls (2009), p. 266). To account for equality in a similar sense, we should make sure that recipients are treated equally according to the rules that apply for them. With respect to autonomous driving, this could be realized by estimating the driving opportunities of all traffic participants in the same way it is done for the autonomous vehicle itself, and a decision is taken by equally respecting their driving needs.

Rawls claims validity of the difference principle for society-wide distribution of social goods, especially in the context of the allocation of wealth and income as consequence of participating in the economic system. For decisions of an autonomous vehicle based on the driving opportunity estimation, only the current driving situation and the assumed nearby traffic participants can be taken into account. This makes it a local distribution problem rather than a society-wide one. Though this is a setting for which the difference prin-

ciple has not been originally developed, we think it is appropriate to transfer it.

The objective of the in-situ decision, based on the difference principle, is that the prospect of the vehicle with the lowest expected driving prospects among all vehicles in the scene is as high as possible. In other words, it will result in driving decisions in which the worst-off vehicles profit the most and it is prevented that high risk situations for certain vehicles are accepted for the benefit of all other vehicles.

Researchers have highlighted the danger of a potential structural discrimination with respect to crash algorithms within autonomous driving (Hin-Yan (2016)). This could happen, when the same or similar crash-principles would be implemented to the majority of autonomous vehicles on the road. According to Hin-Yan, it is very likely that there will be biases in the principled decisions which are multiplied when the same principles are broadly applied (Hin-Yan (2016), p. 164). Discrimination concerns could also be relevant with respect to driving decisions beyond hazard scenarios. For instance, it is a systematic discrimination when certain traffic participants have a general higher likelihood to get in a situation of risk caused by a decision policy of the majority. This is also true, when they are more often hindered in fulfilling their utility or comfort needs.¹⁷ Applying decision rules which take the needs of others equally into account and deciding in favor of the worst-off, rather than solely targeting on self-optimization, will reduce the possibilities of discrimination.

In the implementation section we will continue with the difference principle. However, the two alternative approaches, introduced above, will be briefly considered in the following paragraphs.

One alternative distributive principle is **strict equality**. The major goal for choosing a behavior is that all traffic participants end up having the same or very similar driving prospects. The conditions for realizing equality are that individuals have both similar needs and the goods which are distributed are suitable for giving the same quantity to every individual. It can be assumed that the basic needs (safety, comfort, utility) are important for all traffic participants and similar enough to deploy the same function for the prediction of driving opportunities. With respect to the second condition, there are concerns that for many in-situ driving decisions, options which have very similar opportunities for all participants are not available. This can be assumed because of the complexity of interrelated driv-

¹⁷ For instance Nyholm and Smith assume that autonomous vehicles will be programmed to optimize their passengers' interests: "optimally safe, fuel-efficient, and travel time-efficient" (Nyholm and Smids (2018), p. 2).

ing needs and the limited space of behavior options in traffic.

One way to react to such problem could be to apply a compensating approach. In such an approach, driving scenes are not considered individually but in conjunction which gives room for mechanisms where disadvantages in one situation are balanced out with benefits in another scene. The challenge is that because constellations of traffic participants are changing quite fast a more global accounting system, similar to what Mladenovic and McPherson are suggesting, may be required (Mladenovic and McPherson (2016)). A vehicle which is thwarted by another vehicle in one scene could be compensated by giving it way (allowing it to progress faster) in another scene. If such a compensation system is appropriate with respect to risk allocation is questionable. A practical downside is that such an infrastructure has high implementation costs.

The **utilitarian principle** could be applied in the behavior selection by choosing the option for which the overall driving opportunities of all traffic participants is in sum the highest. This concept can be criticized by referring to a concern generally brought up for utilitarian approaches. In its pure understanding it does not prevent distributive outcomes in which single persons or groups are, maybe systematically, discriminated. An increase of utility can be realized by supporting the welfare of all recipients likewise or by supporting the majority even more and leaving some people or a minority group behind. Applying the utilitarian principle could result in situations where options are chosen for which the majority receive a high utility but the minority, for instance one worst-off traffic participant, is confronted with a relatively dangerous driving situation.

In the following section we will focus on the difference principle and show how this approach can be implemented in an autonomous driving system.

6 Implementation

To illustrate the practical impact of our proposed concept, we apply distributive justice to a driver assistant system for partially autonomous highway driving.¹⁸ Designed to provide improved comfort for everyday driving, this allows us to test our concepts in particular in the non-hazard context.

An overview flow diagram of this system can be found in Fig. 1 (black parts). It receives perception

¹⁸ The assistance system is an extension of an adaptive cruise control system with behavior prediction capabilities as introduced in Kleinehagenbrock et al (2015), Schmuelderich et al (2015).

data from the vehicle sensors and uses the information about traffic participants and lane infrastructure to construct a 360° scene representation centered around the autonomous vehicle. The core of the system is the behavior planning module¹⁹, which uses this representation to predict likely evolutions of the traffic scene in the future, and then optimizes up to three vehicle behavior trajectories (one each for straight driving, and lane-change left and right) for each future scene which balances costs and benefits for the autonomous vehicle. The future scene is constructed using predictions of all possible next behaviors of the other vehicles, that means all behaviors which are not stated as very unlikely (more than 10 % probability), based on their current driving properties and behavior history.

Additionally, these predictions are done separately for each possible behavior of the autonomous vehicle, that is we compute how the behavior probabilities would change, if it would select a certain maneuver. Trajectories are generated by adjusting parameters of longitudinal and lateral motion representations along a cost gradient, for example as velocity-profile and the timing for a lane-change. The costs represent aspects of driver comfort, by penalizing hard braking and acceleration maneuvers and frequent lane-changes, safety, by rewarding sufficient space to other traffic participants, and utility, through the incorporation of a desired speed (for an overview, see Tab. 2).

Cost Type	Driving Needs
time-gap	safety + comfort
braking	safety + comfort
acceleration change (jerk)	comfort
lane-change	comfort
preference-speed	comfort + utility

Table 2 The table shows the cost terms which represent the driving opportunities. Every cost type has an underlying function parametrized according to regular human driving data, for example an exponential increase of costs when the gap to other participants (time-gap) becomes very small

Based on the results of behavior planning a selection mechanism chooses the trajectory that should be communicated to the car’s actuation system. The original algorithm for this, which we will call *ego behavior selection*, selects the trajectory which has the lowest cost for the future situation with the highest probability.

Alternatively, we use an algorithm implementing our distributive justice concept, which we call *fair behavior selection*. It is required to predict the costs of

¹⁹ For technical details on the system, see Weisswange et al (2018).

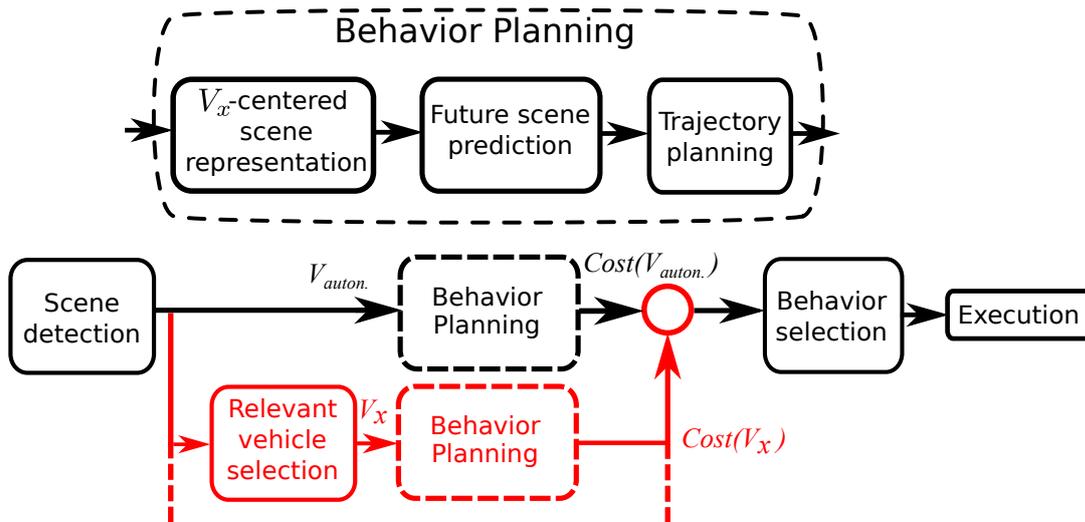


Fig. 1 The figure shows the processing flow of the driver assistance system. Black parts are shared between *ego* and *fair* behavior selection, red modules are only used in the latter. For the *fair* system, the planning module is run for every vehicle $V_x \in \{V_1, \dots, V_n\}$ in the surrounding and behavior selection is done using all individual costs, whereas the basic variant only considers the cost of the own trajectory ($V_x = V_{auton.}$)

the other traffic participants (driving opportunities) in a similar way to make driving quality and safety needs comparable. For that we use basically the same original behavior planning module also for the other vehicles. The currently possible autonomous vehicle behavior options are used to fix part of the future traffic scene. For every vehicle that is affected by at least one of the autonomous vehicle's behavior options, their path costs are predicted. This is done through a full behavior planning loop that applies the same optimization mechanisms used for the autonomous vehicle planning. For the first four cost terms in Tab. 2 we used the same parametrization for all traffic participants. The preferred speed of the other drivers is not directly assessable, we therefore approximated it by the maximum velocity of the vehicle that was perceived so far. It is not very likely that the vehicles exactly follow this optimal trajectory, however, it provides us with a good estimate of relative driving costs between the traffic participants. This processing is repeated for every relevant vehicle in the influencing sphere of the autonomous vehicle. Subsequently, the costs of all traffic participants are taken into account for the behavior selection of the autonomous vehicle. As can be seen by the red icons in Fig. 1, we only have to adapt a small part of the original system for this.

In the next section we will describe how this processing flow represents the aspects of a just distribution of road goods.

6.1 With Regard to the Four Dimensions of Distributive Justice

The road environment as well as limitations of sensing range and algorithmic processes, require a discussion of the mapping between our general distributive justice concept for traffic and the concrete application in an assistance system.

The distributive good, as we have argued before, is the public road with its objective and conditions of usage for instance comfort, safety (inverse: risk) and utility.

Because our system can only perceive traffic participants in the close vicinity of the vehicle (maximum up to 200 m to front and back), we can only consider the distribution of road usage on a *local* and *short-term* scale. However, since our control parameter for providing a just distribution is the behavior of the autonomous vehicle, this actually covers to a very large degree the influencing sphere of our system. Accordingly, the recipients of the distribution will be limited to those traffic participants (respectively their drivers/passengers) which interact directly with the autonomous vehicle on the local road patch. The needs of a traffic participant are perfectly satisfied when inhabitants reach a destination in the preferred timeframe, with minimal risk under conditions matching his or her preferred style (comfort). We can therefore apply the same function that is used for the autonomous vehicle trajectory planning to assess the cost that would be induced in any other traffic participant.

There exist cases, where you could argue for specific characteristics of vehicles, that could privilege them in the distribution of costs.²⁰ One could also think about including the number of people in a car, or the vulnerability of a vehicle to bias the assignment of the goods. However, we did not consider those cases, because our sensing systems (and most current production sensors) are not reliably able to detect these special characteristics. It is anyway a topic of ethics discussions, if the number of people or other factors should be taken into account at all.²¹ From the perspective of the driver or passenger of an autonomous vehicle it might also be valid to argue for a special role for himself, as particularly discussed for accident decisions. It could be argued for a biased assignment based on the self-protection right of humans. With respect to that, self-sacrifices of individuals for the sake of others can under no circumstances be demanded. In the literature, it is under discussion, if and how the self-protection right applies for accident cases.²² It has to be examined even further if in situations where no serious harm is immediate self-protection rights are legitimate factors which have to be considered.

We decided to use the equality principle and weight all participant's costs with the same factor. For the rule of distribution we used the difference principle, implemented through the selection of the autonomous vehicle behavior which is associated with the minimum cost value for the traffic participant with the maximum cost given this behavior.

In short, our mapping of the four dimensions of distributive justice are:

1. Distributive good: usage of the current part of the public road, which allows a fulfillment of comfort, safety and utility needs, inversely represented through the cost-function.
2. Recipients of the distribution: all actors (traffic participants are equivalent to the drivers) in the influencing sphere of the autonomous vehicle's actions (first-order)
3. Characteristics of recipients: no special role identified – equality
4. Rule of distribution: Minimize maximum cost among all recipients (difference principle)

The next section will present the differences in behavioral outcome of a self-interested decision making

²⁰ For instances an ambulance could be a privileged traffic participant.

²¹ see for example: Ethics-Commission (2017).

²² in Ethics-Commission (2017): Paragraph 9 and in more detail in the appendix (only in the German version).

versus the fairness-centered (distributive justice) approach in an application to real driving data.

7 Results

For the evaluation of our implementation on the partially autonomous decision system, we used data from two real-world highway drives with our experimental vehicle platform. We have selected two typical driving situations for which we can best illustrate the fairness effect. The first situation-type we call *Cut-In* and the second *Fast Successor*.

Cut-In situations appear when the autonomous vehicle is on one of the left lanes, while another vehicle drives to the right in front of it and approaches a slower vehicle, for example a truck. Fig. 2 illustrates this type of situation. In the example the autonomous vehicle (V1 in Fig. 2) has two useful behavior options, a lane-change maneuver targeting the left lane and continuing to drive straight. At the same time, the scene prediction module will provide two probable hypotheses for a future layout, one where the right car (V2 in Fig. 2) will cut into the lane of the autonomous vehicle, and one where it will stay behind the truck until we have passed. Depending on the autonomous vehicle behavior, the probabilities for these alternatives will change. If the autonomous vehicle would do a lane-change, the available gap for V2 would be much larger, and this would lead to a higher cut-in probability. In the concrete situation shown here, only two of the four combination reach a significant probability (i.e. the vehicle is not expected to change lane if the autonomous vehicle drives straight). Therefore the planner calculates optimal autonomous vehicle trajectories only for the two dominant combinations shown in Fig. 2 bottom (color coded).

Since there is currently no vehicle behind the autonomous vehicle, the only vehicle that will be directly affected by its behavior is vehicle V2. To take its driving needs into account, we also compute its optimal trajectory for each of the autonomous vehicle behavior options. The two available future situations are V1 changes left and V2 changes left, and V1 drives straight and V2 drives straight.

The evaluation of the situation has shown that the predicted costs of a vehicle approaching a slower vehicle in front increase because it has to brake. The autonomous vehicle has to do the lane-change to the left before it becomes too dangerous for vehicle V2 to do the conditional lane-change.

We will explain how the implementation of the ethical decision rule differs from solely self-interest-driven

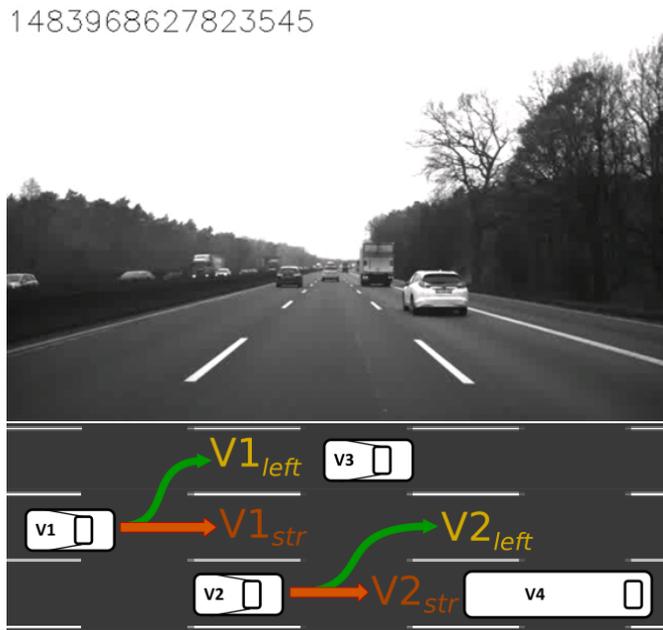


Fig. 2 Top: Screenshot of a *Cut-In* scene. Bottom: Birds-eye-view representation of the traffic situation and the behavior options for the autonomous vehicle V1 and the relevant other traffic participant V2

decision principles. By taking the costs of vehicle V2 into account, the min-max principle of the fairness approach would more likely lead to a lane-change of the autonomous vehicle. A lane-change behavior is executed when the maximum costs of any of the relevant vehicles with respect to the lane-change behavior is lower (min-condition) than for the alternative option (continuing straight). Considering the costs of vehicle V2, caused by approaching the slower truck in the fairness approach leads to a lane-change decision. In contrast, following the *ego behavior selection* approach, solely the costs of the autonomous vehicle’s behavior options would be compared: Going straight is preferred (straight driving has more comfort) even when lane-change would be relatively cheap option as well (i.e. because of an empty left lane).

A second scene-type, in which fair behaviors can have a significant impact is the *Fast Successor* situation. It is characterized by the autonomous vehicle doing a lane-change to open the lane for a vehicle approaching from behind with a higher speed. One such situation is shown as example in Fig. 3. The autonomous vehicle (V1) is overtaking a truck, while a fast vehicle is approaching from behind. Giving way by using the next gap on the right is reducing the load to brake for vehicle V2, but requires the autonomous vehicle to slow down to fit in the gap between the truck (V3) and vehicle V4. Applying the difference principle, the

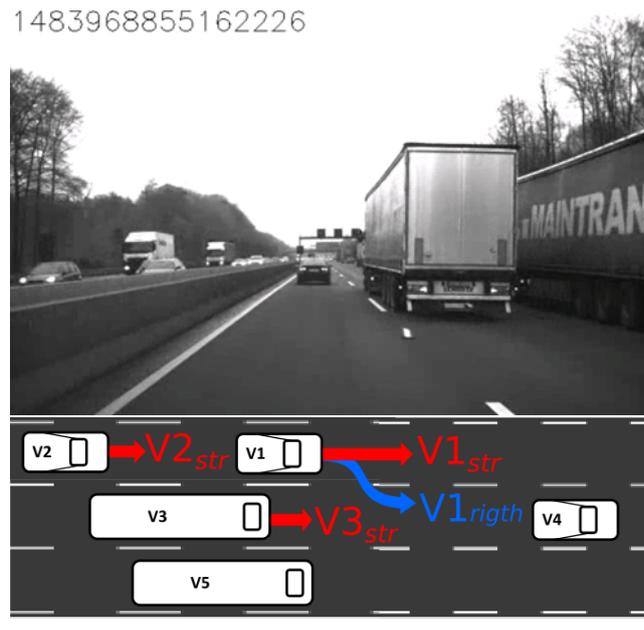


Fig. 3 Top: Screenshot of a *Fast Successor* scene. Bottom: Birds-eye-view representation of the traffic situation and the behavior options for the autonomous vehicle V1 and the relevant other traffic participants V2 and V3

behavior is selected which leads to the smallest cost for the vehicle suffering maximum costs in this situation. If the costs for braking for V2 are very high, the autonomous vehicle is requested to brake to fit in the gap to let vehicle V2 pass. A behavior selection which would be solely based on the self-interest of the autonomous vehicle would choose a straight driving behavior because the right lane-change has higher costs than going straight.

Both examples illustrate how the implemented difference principle can have effect on behavior decisions on a public highway – based on real data.

8 Conclusion

In this section, we summarize the results of the interdisciplinary approach and briefly discuss possible next steps. We explained why ethical concepts should be implemented for autonomous vehicles, not only to allow vehicles to cope with hazard situations, but rather to ethically align all driving behavior in public traffic.

An ethical approach beyond hazard situations is of relevance because the road (seen as public good) is a shared operational space where most decisions of traffic participants impact the behavioral possibilities of others. We discussed in the paper why and how autonomous vehicle should be designed respecting the ethical significance of the decisions they will make.

With respect to the *how-to*, we argued that distributive justice can serve as ethical framework which should be applied to coordinate operations of autonomous vehicles in traffic space. In order to transfer distributive justice into the traffic context, we reasoned that all traffic participants in the in-situ driving scene could be seen as equal recipients for distribution with similar driving needs (safety, comfort and utility). Regarding the concrete rule for aligning the behavior of autonomous vehicles, we showed that the *difference principle* is appropriate to be applied here. It leads to operational decisions which are in favor of the participants with the worst-off driving prospects in a traffic scene.

Furthermore, we argued that distributive justice as decision principle should be applied for all autonomous vehicles. In contrast, in the scope of only-human driving and with respect to the scenes we have described in the implementation section, it is not seen as mandatory for humans to drive in a certain way, at least it is not explicitly required by traffic law.²³ Unlike human drivers who are rightly allowed to act within a certain behavioral freedom whereby the results lie within their responsibility, autonomous vehicles are not free actors in a human sense. They can be rather characterized as actors able to perform a domain-specific principled agency. As the behavior possibilities of such agents are aligned by explicit principles, we think it lies in the designers' forward-looking responsibility to specify them in a way that takes the described normative significance of decisions into account. We think distributive justice serves as appropriate framework to fulfill this requirement.

Based on the theoretical justification, the implementation part has shown how the concepts of just distribution becomes effective in a realistic traffic scenario. The core element of the implementation to align with the ethical principle is the planning module which ensures the equality of treatment between all participants and the min-max optimization, which realizes the difference principle. The results show that an implementation of an explicit ethical principle is feasible for a highway scenario and can produce reasonable driving decisions.

²³ Street regulations for example in Germany may contain general traffic principles, which are relevant for such scenes. The street regulation in Germany obliges cautious driving and mutual respect and understanding as a first principle (Paragraph (1) in the StVO - Germany). It is a vague formulation but still an important principle because it allows people to be held partly responsible for e.g. being incautious, even when strict traffic law (for example the right of way) is on their side - Paragraph (1) in the StVO - Germany (http://www.gesetze-im-internet.de/stvo_2013/).

The exact parameters which should be taken into account and the weighting between them are both a topic for a broader social and ethical discussion and conditioned by the technical feasibility which is constrained by e.g. sensing and detection capabilities. The four dimensions of just distribution can also be seen as methodological framework for the future discussion of relevant factors.

Discussing principles of distributive justice are best thought of as a base for "providing moral guidance" for social institutions, which can be the legal system or political or public constitutions (see: introduction of Lamont and Favor (2017)). The concepts of distributive justice which are applied for non-hazardous autonomous driving should be make use of by responsible designers or manufactures. The approach, we have shown here, is not dependent on a *global* traffic infrastructure, which means it does not require any technical measures beyond the ones a vehicle can be equipped with. It is straightforward, that a certain car-to-car communication infrastructure could help to move forward from a hypothetical planning from a third-person perspective to a first-person planning (every vehicle for itself) and a real-time sharing of trajectory costs for different options for a globally coordinated behavior planning.

Acknowledgments

The authors would like to thank the anonymous reviewers for their support towards improving the overall quality of the paper through the process.

References

- Anderson ES (1999) What is the point of equality? Ethics 109(2):287–337
- Bentham J (1996) The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation. Clarendon Press, Oxford
- Bojarski M, Del Testa D, Dworakowski D, Firner B, Flepp B, Goyal P, Jackel LD, Monfort M, Muller U, Zhang J, Zhang X, Zhao J, Zieba K (2016) End to End Learning for Self-Driving Cars. In: Advances in Neural Information Processing Systems 29 (NIPS 2016) Deep Learning Symposium, Barcelona, Spain, URL <http://arxiv.org/abs/1604.07316>
- Bonnefon JF, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. Science 352(6293):1573–1576
- Charisi V, Dennis LA, Fisher M, Lieck R, Matthias A, Slavkovik M, Sombetzki J, Winfield AFT, Yampol-

- skiy R (2017) Towards moral autonomous systems. CoRR abs/1703.04741, 1703.04741
- Coeckelbergh M (2016) Responsibility and the moral phenomenology of using self-driving cars. *Appl Artif Intell* 30(8):748–757
- Cohen GA (1989) On the currency of egalitarian justice. *Ethics* 99(4):906–944
- Danaher J (2016) Robots, law and the retribution gap. *Ethics and Information Technology* 18(4):299–309
- Dworkin R (1981) What is equality? part 2: Equality of resources. *Philosophy & Public Affairs* 10(4):283–345
- Elster J (1993) *Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens*. G - Reference, Information and Interdisciplinary Subjects Series, Russell Sage Foundation, New York
- Ethics-Commission (2017) Automated and connected driving. <https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf>, accessed: 2018-01-30
- Ferretti MP (2016) Risk imposition and freedom. *Politics, Philosophy & Economics* 15(3):261–279
- Finnis J (2011) *Natural Law and Natural Rights*. Clarendon Law Series, Oxford University Press, Oxford
- Frankena WK (1963) *Ethics*. Prentice-Hall Foundations of Philosophy Series, Prentice-Hall Incorporated, New Jersey
- Gerdes JC, Thornton SM (2016) Implementable ethics for autonomous vehicles. In: Maurer M, Gerdes JC, Lenz B, Winner H (eds) *Autonomous Driving: Technical, Legal and Social Aspects*, Springer, Berlin, Heidelberg, pp 87–102
- Goodall N (2014a) Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board* 1(2424):58–65
- Goodall N (2014b) Machine ethics and automated vehicles. In: Meyer S Gereon; Beiker (ed) *Road Vehicle Automation*, Springer International Publishing, Cham, pp 93–102
- Harsanyi JC (1977) Morality and the theory of rational behavior. *Social Research* 44(4):623–656
- Hevelke A, Nida-Rümelin J (2015) Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics* 21(3):619–630
- Hin-Yan L (2016) Structural discrimination and autonomous vehicles: Immunity devices, trump cards and crash optimisation. *What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016* 290:164–173
- Jenkins R (2017) The need for moral algorithms in autonomous vehicles. In: Otto P (ed) *3THICS: A Reinvention of Ethics in the Digital Age?*, iRights Media, Berlin, pp 88–97
- Keeling G (2017) Commentary: 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* 11:247
- Kleinehagenbrock M, Nishigaki M, Kastner R, Schmuëdderich J, Rebhan S, Weisswange TH, Kamiya H, Mori N, Kusahara S, Ishida S (2015) Introduction of intelligent adaptive cruise control (i-acc): A predictive safety system. In: *FAST-zero'15: 3rd International Symposium on Future Active Safety Technology Toward Zero Traffic Accidents*, 2015
- Lamont J, Favor C (2017) Distributive justice. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, winter 2017 edn, Metaphysics Research Lab, Stanford University
- Leben D (2017) A rawlsian algorithm for autonomous vehicles. *Ethics and Information Technology* 19(2):107–115
- Lin P (2015) Why ethics matters for autonomous cars. In: Maurer M, Gerdes JC, Lenz B, Winner H (eds) *Autonomous Driving: Technical, Legal and Social Aspects*, Springer, Berlin, Heidelberg, pp 69–85
- Luetge C (2017) The german ethics code for automated and connected driving. *Philosophy & Technology* 30(4):547–558
- Lütge C, Rusch H, Uhl M, Luetge C (2014) *Experimental ethics: Toward an empirical moral philosophy*. Springer, Berlin, Heidelberg
- Margalit A (1998) *The Decent Society*. Harvard University Press, Cambridge, Massachusetts
- Mill JS (1861) *Utilitarianism*. Oxford University Press, Oxford
- Misselhorn C (2015) Collective agency and cooperation in natural and artificial systems. In: *Collective Agency and Cooperation in Natural and Artificial Systems*, Springer, Berlin, Heidelberg, pp 3–24
- Mladenovic MN, McPherson T (2016) Engineering social justice into traffic control for self-driving vehicles? *Science and Engineering Ethics* 22(4):1131–1149
- Nussbaum MC (1992) Human functioning and social justice: In defense of aristotelian essentialism. *Political Theory* 20(2):202–246
- Nyholm S (2018a) Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci. *Science and Engineering Ethics* 24(4):1201–1219
- Nyholm S (2018b) The ethics of crashes with self-driving cars: A roadmap, ii. *Philosophy Compass* 13(7):1–10

- Nyholm S, Smids J (2016) The ethics of accident-algorithms for self-driving cars: an applied trolley problem? *Ethical Theory and Moral Practice* 19(5):1275–1289
- Nyholm S, Smids J (2018) Automated cars meet human drivers: responsible human-robot coordination and the ethics of mixed traffic. *Ethics and Information Technology*
- Persson I (2007) A defence of extreme egalitarianism. In: Holtug N, Lippert-Rasmussen K (eds) *Egalitarianism: New Essays on the Nature and Value of Equality*, Clarendon Press, pp 83–98
- Rawls J (2009) *A Theory of Justice: Revised edition*. Harvard University Press, Cambridge, Massachusetts
- Schmidt VH (1992) Adaptive justice: Local distributive justice in sociological perspective. *Theory and Society* 21(6):789–816
- Schmuedderich J, Rebhan S, Weisswange TH, Kleinhagenbrock M, Kastner R, Nishigaki M, Kamiya H, Mori N, Kusuhara S, Ishida S (2015) A novel approach to driver behavior prediction using scene context and physical evidence for intelligent adaptive cruise control (i-acc). In: *FAST-zero'15: 3rd International Symposium on Future Active Safety Technology Toward Zero Traffic Accidents, 2015*
- Shafer-Landau R (2007) *Ethical Theory: An Anthology*. Blackwell Philosophy Anthologies, Wiley-Blackwell, West Sussex
- Sparrow R (2007) Killer robots. *Journal of Applied Philosophy* 24(1):62–77
- Waymo (2017) Waymo safety report: On the road to fully self-driving. <https://waymo.com/safetyreport>, accessed: 2018-03-28
- Weisswange TH, Rebhan S, Bolder B, Steinhardt NA, Joublin F, Schmuedderich J, Goerick C (2018) Intelligent Traffic Flow Assist: Optimized Highway Driving Using Conditional Behavior Prediction. *IEEE Intelligent Transportation Systems Magazine* submitted