Limits and Prospects of Ethics in the Context of Law and Society by the Example of Accident Algorithms of Autonomous Driving

Peter Klimczak

1. Autonomous Driving

At the beginning of 2014, the Society of Automotive Engineers (SAE) published the SAE J3016 standard, which classifies road vehicles, i.e. automobiles, in terms of their automatic functions. SAE J3016 specifies six levels, which – despite a number of alternative classifications, e.g. by ADAC or the German Federal Highway Research Institute ("Autonomes Fahren" 2021) – have since become established even in Germany ("Autonomous Driving" 2018):

- *Level 0: No Automation* – Zero autonomy; the driver performs all the driving, but the vehicle can aid with blind spot detection, forward collision warnings and lane departure warnings.
- *Level 1: Driver Assistance* – The vehicle may have some active driving assist features, but the driver is still in charge. Such assist features available in today's vehicles include adaptive cruise control, automatic emergency braking and lane keeping.
- *Level 2: Partial Automation* – The driver still must be alert and monitor the environment at all times, but driving assist features that control acceleration, braking and steering may work together in unison so the driver does not need to provide any input in certain situations. Such automated functions available today include self-parking and traffic jam assist (stop-and-go traffic driving).
- *Level 3: Conditional Automation* – The vehicle can itself perform all aspects of the driving task under some circumstances, but the human driver must

always be ready to take control at all times within a specified notice period. In all other circumstances, the human performs the driving.

- *Level 4: High Automation* – This is a self-driving vehicle. But it still has a driver's seat and all the regular controls. Though the vehicle can drive and "see" all on its own, circumstances such as geographic area, road conditions or local laws might require the person in the driver's seat to take over.
- *Level 5: Full Automation* – The vehicle is capable of performing all driving functions under all environmental conditions and can operate without humans inside. The human occupants are passengers and need never be involved in driving. A steering wheel is optional in this vehicle.

Until five years ago, there was real euphoria around the development of autonomous cars, but since then developers, established car manufacturers and, even more so, IT companies, above all Google, have become disillusioned. John Krafcik, the CEO of Waymo and, as such, of the subsidiary of Google or rather Alphabet, to which Google outsourced its efforts in the field of self-driving cars in 2017, recently stated that the implementation of autonomous driving is "an extraordinary grind [and] a bigger challenge than launching a rocket and putting it in orbit around the earth" ("Rolling out" 2021).

In fact, we are still at most at level 3 of the SAE standard. The vehicles above this level are still being developed or tested as prototypes and are ready for use only to a very limited extent (on the current state of the art of autonomous vehicles, cf. Kleinschmidt/Wagner 2020: 7). The reasons for the relatively slow progress are of a technical nature: while driving on highways can be machine-sensed and processed sufficiently thanks to artificial intelligence, mixed traffic (cars, cyclists, pedestrians) on rural roads and in residential areas (especially cities) poses a problem that will remain difficult to solve in the future due to the multimodal, polysemous and culturally differentiated communication of the various road users (Färber 2015: 127). While technicians in particular, and linguists and semioticians more marginally, are concerned with these problems, ethicists and lawyers focus on other interrelated problems that are emerging.

The problems being discussed in the areas of ethics and jurisprudence have not yet become acute for manufacturers due to the fact that, currently, neither highly nor fully automated vehicles regularly participate in traffic. However, these ethical and legal questions eventually will arise and will have

to be answered by the vehicle manufacturers and third-party providers of hardware and software for autonomous driving – such as Waymo, which ultimately wants to transfer the Android business model from the smartphone sector to the automotive sector[1] –, if the manufacturers do not want to make their product unattractive for customers and thus also investors due to moral sensitivities or real safety interests, or if they want to protect themselves or their own developers from civil and especially criminal prosecution.

Although technical development can continue in isolation from such questions, any failure to answer these ethical and legal questions will inhibit operation and sales and – since technical innovation[2] essentially involves, not least, the offering and selling of a new, innovative product – will prove inhibiting to technical innovation.[3] In this respect, ethics and law should, can and must be regarded as determinants of technical innovation.

1 The business model, namely, of not producing the device, i.e. the smartphone or the car, but 'merely' offering the smartphone operating system or the hardware and software components for autonomous driving and thus achieving market dominance independent of the respective manufacturer.

2 In my opinion, a successful definition of innovation has been proposed by Hoffmann-Riem, for whom innovations are – to put it simply – "significant, namely practically consequential improvements" that "can be socially desirable or undesirable," whereby the former is the case "if the innovation – for instance in the form of new products, processes or institutions – can contribute to overcoming a problem that is seen as in need of a solution" (Hoffmann-Riem 2016: 12).

3 In the legal context, these are then legally uncertain conditions. Legal uncertainty – according to a central tenet of political economics – prevents the growth and stability of an economy, since the actors cannot assess the negative legal consequences of their actions (Wagner 1997: 228ff.). That this causal relationship can be transferred from macroeconomics to microeconomics, especially with regard to the realization of innovative technologies, is shown case-specifically in two contributions to the recent volume "Recht als Infrastruktur für Innovationen" (Datta 2019; Servatius 2019). Two older, yet still topical contributions appear in the anthology "Innovationsfördernde Regulierung"; of these, one implicitly (Kühling 2009: 62) and the other explicitly (Roßnagel 2009: 336) addresses said causality. Interestingly – or regrettably – the established introductions and handbooks to technology and innovation management do not address this relationship. See, for example, Abele 2019; Möhrle/Isenmann 2017; Wördenweber et al. 2020; Albers/Gassmann 2005.

2. Accident Algorithms – Law and Ethics in Their (Respective) Discourses

Ethical and legal questions and challenges with regard to autonomous driving are discussed in particular in the context of so-called accident algorithms and these, in turn and taken to the extreme, in the context of so-called trolley problems. Feldle, whose thesis on "Unfallalgorithmen. Dilemmata im automatisierten Straßenverkehr" is comprehensively dedicated to the topic, writes:

> In an unforeseen situation, where a human would only reflexively jerk the wheel around, a technical system can make pragmatic decisions. But even sophisticated "algorithm-driven collision avoidance systems" will not render all accidents avoidable. In extreme scenarios, if multiple interests are threatened, a technical system could in the future determine which legal protected interest is sacrificed for the good of another. Fatal accidents will also be part of this assessment. (Feldle 2018: 22)[4]

The latter in particular, Feldle goes on in the words of Schuster, is "one of the last largely unresolved legal problems of automated road traffic" (Schuster 2017: 13) and he refers to Hörnle's and Wohler's programmatic statement: "We are still only at the beginning of the necessary normative discussion" (Hörnle/Wohlers 2018: 12f.).

Legal treatises deal with ethical issues only marginally and mostly only in the context of noting that ethics has raised the aforementioned trolley problem.[5] A positive exception in this respect is the thesis cited above by Feldle from 2018, which, for example, deals with the position or rather proposals of the two ethicists Helvelke and Nida-Rümelin and subjects them to a legal examination (Feldle 2018: 189ff.). However, even Feldle does not provide an in-depth discussion of the ethical concepts. Conversely, ethical treatises on the trolley problem hardly address the connection between ethics and law. The two monographs by Zoglauer (1997 and 2017) are a welcome exception in this regard. The discussion presented in the following, however,

4 Quotations from German sources are translations.

5 Which – as will be seen later – is only half true, since such scenarios, as far as can be reconstructed, were first developed in jurisprudence and then popularized by ethics, with the result that they have found their way back into legal discourse.

differs from Zoglauer's in that, on the one hand, it explicitly deals with autonomous driving and, on the other hand, it does not regard the problems of legal derogation and judicial or supreme court deliberation, which Zoglauer has clearly identified, as a problem, but as a solution – namely as a solution for technical innovations.

It must be clearly stated that I assume that contradictory regulations have a negative effect on technical innovation,[6] which is why risks, and not opportunities, are always in the foreground. If one wishes to and shares the premises and arguments of the following remarks, the more implicit, less explicit call to create regulations can be seen as an opportunity, since, on the one hand, these create (legal) security for technical innovation and, on the other hand, corrections or, formulated more cautiously, changes will continue to be possible at any time thanks to the discursive character of ethics and law. By contrast, non-regulations offer neither security nor, by their very nature, the possibility of correction, since the latter would first require something that could be corrected.

One can of course take the position that non-regulations and absolute freedoms would be conducive, even indispensable, for technical innovation.[7] However, I would counter that in the case of accident algorithms, and especially in the case of accident algorithms in the context of the inevitable endangerment of human life, an extremely sensitive area, both ethically and legally, is involved and undesirable developments or an "anything goes" approach in such sensitive areas leads to a dramatic reduction in the trust and acceptance of society and consumers for a technical innovation (Klimczak et al. 2019: 44f.). Moreover, we are no longer in the initial stages of this technical development, in which prohibitions on thinking can indeed have a negative impact on technical innovation (Hoffmann-Riem 2016: 262ff.), but rather on the threshold of realizing the technology at issue.

6 If a contradiction cannot be resolved (see below, in particular with regard to the derogation principles), the situation would again be one of legal uncertainty. On legal uncertainty, see the remarks in the footnotes above.

7 This view was especially true in the 1980s and 1990s. Cf. Hoffmann-Riem 2016: 26f.

3. Human and Machine Decisions about Life and Death

For (almost) all of the following reflections on autonomous driving, a situation is assumed – as already mentioned – in which an accident with fatal personal injury has become unavoidable. One possible variation on such a scenario would be that an autonomously driving vehicle is driving on a country road. Two other vehicles are coming towards it, due, for example, to an unsuccessful overtaking maneuver. In the car that is in the same lane as the autonomously driving car, i.e. in the passing car, there are three occupants, while there is only one person in the other lane, i.e. in the car being overtaken. The only choices the vehicle AI has with regard to steering (i.e. longitudinal movement) are to keep to its lane, i.e. to make no changes, or to change lanes, i.e. to behave actively. In the first case (lane keeping), three people are killed, in the second case (lane changing), one person is killed.[8]

Examples like these are adaptations of a popular thought experiment in ethics that was introduced by Karl Engisch from a criminal law perspective[9] in his habilitation thesis in Law in 1930, and which Philippa Foot reformulated in 1967,[10] this time in English and in a moral philosophical context (as a comparative case for the abortion issue that was her actual focus); this experiment subsequently became extremely well-known (with active participation by Foot) as the trolley problem. Owing to the chronology, the thought experiments of Engisch, Foot and others do not in any way involve autonomous vehicles, and so the central difference between these classical case descriptions and those adapted for autonomous driving is between the different deciding agents. Therefore, the question that should be pursued by means of the original thought experiments is as to which premises, which presup-

8 For similar scenarios in the context of autonomous driving, see the list of common types in Schäffner 2020: 29f.

9 Cf. Engisch 1930: 288. Verbatim (in translation): "It may be that a switchman, in order to prevent an imminent collision which will in all probability cost a large number of human lives, directs the train in such a way that human lives are also put at risk, but very much less so than if he were to let things take their course."

10 Verbatim: "To make the parallel as close as possible it may rather be supposed that he is the driver of a runaway streetcar which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed" (Foot 1967: 8).

positions form the basis when people make moral decisions:[11] as a driver, do I steer into the car in the oncoming lane because I see only one occupant there and spare the three people in the car that is in my lane? Or would I rather collide with the three occupants, i.e. accept more deaths, because in this case I do not have to actively intervene in the steering and the choice is therefore made by omission?

Such questions – and this is the unintended point of the original thought experiment – are of a completely counterfactual nature (Hevelke/Nida-Rümelin 2015: 8): In a real-case scenario, the driver would not make a considered moral decision, but an instinctive one (Dilich et al. 2002: 238).[12] The rationality of human decision-making is limited even in such situations that do not take place under emotional and temporal pressure, because people have access to only limited information in decision-making situations and their capacities to process it as well as to consider the consequences of different options are limited (Klimczak et al. 2019: 41): people, therefore, fundamentally decide all too often on the basis of intuition and emotions without careful analysis, which is why human decisions are always susceptible to different types of cognitive distortions.

In contrast to human decision-making, machine decision-making is necessarily based on a calculation and thus an assessment.[13] However, this calculation is determined in advance: fundamental decisions must already be made during the development of the AI in question. This applies not only in the case of 'classical' AI programming due to the algorithms that need to be written, but also in the case of so-called machine learning procedures. For the latter, the selection and processing of training data is also added as a further decision. The machine decision is then the result of a calculation and only of a calculation. Thus, an automated decision process is free of cognitive distortions, but in general, and in the case of machine learning in particular, not free of 'biases' (Klimczak et al. 2019: 41): for example,

11 Accordingly, Foot immediately follows on from the quote above: "The question is why we should say, without hesitation, that the driver should steer for the less occupied track, while most of us would be appalled at the idea that the innocent man could be framed" (1967: 8).

12 The fact that people normally have no difficulty in explaining their decisions a posteriori/ex post does not change the fact that this decision did not actually come about in this way. It is a constructed, verbal rationality, not a factual one. Cf. Klimczak et al. 2019: 39.

13 Cf. here and in the following: Berkemer/Grottke 2023.

a character recognizer (OCR) expects to recognize the letter E much more readily than the letter X, which is justified rationally by the much higher frequency of the former, the so-called a priori probability, in the written German language. However, precisely this justified bias can lead to wrong decisions in individual cases. For this reason, the fact that migrants are underrepresented in management positions also leads to the fact that an algorithm based on a corresponding data set will hire fewer migrants.[14]

Since such objectively wrong decisions are grounded in the data set, AI applications based on machine learning principles nevertheless make exclusively rational decisions in the sense of both practical rationality – oriented to the relationship between ends and means – and formal rationality – i.e. based on universal rules.[15] For Artificial Intelligence, there is no room for spontaneity, no possibility of deviation from the calculations. This is true, even if popular and press opinion say otherwise, with regard to every AI procedure, hence also with regard to the currently fashionable machine learning, e.g. by means of neuronal networks. The difference between the former and the latter methods is not that the processes of neural networks are not completely computable (they absolutely are), but that they are neither formal-semantically nor intuitively interpretable, since the immense number of neuron connections is utterly incomprehensible for humans, which is why the former method is also counted among the white box methods, the latter among the black box methods (which also include hidden Markov models) (Klimczak et al. 2019: 40f.).

This means, on the one hand, that humans explicitly and implicitly (the latter through the provision of training data in the case of machine learning procedures) define criteria according to which decisions are then made by the AI in specific cases. On the other hand, this means that the purely counterfactual question in the original thought experiment about the basis of each moral decision and thus about the ethical legitimacy or justifiability of this decision basis now actually becomes relevant and virulent.

14 Pattern recognition, which inherently underlies machine learning, thus develops a performative effect that not only reproduces but also naturalizes detected similarities and differences. Cf. Chun 2018.

15 On these and other types of rationality following Max Weber, see Kalberg 1980.

4. Utilitarianism or Consequentialism

An oft-discussed approach in the technology discourse to answering the question of which of the two oncoming cars the autonomous driving vehicle should collide with is so-called utilitarianism. The goal of the utilitarian approach is the mathematical optimization of the benefit and harm consequences of an action (Zoglauer 1997: 161). Since it can be operationalized by machines, such an approach seems not only technically feasible,[16] but, at first glance, also rational, precisely because it can be operationalized by machines. While the former, i.e. technical feasibility, seems to be promising for the development of AI algorithms and thus technical innovations, the latter, i.e. rationality, seems to justify the utilitarian approach from the very outset in a society in which rationality is one (if not the most important) legitimation criterion: 'modern' societies are, after all, characterized not least by the fact that they are permeated by expectations of rationality. Accordingly, rule-governed decisions made on the basis of algorithms are evidence of the highest rationality (Meyer/Jepperson 2000; Lemke 2000).[17]

The problems in the concrete application or operationalization of utilitarian approaches, however, start with what shall be considered harm and what shall be considered benefit and how both are to be quantified. Let us first turn to the simpler question: the consequences of harm. In the context of the above example, it was assumed that the killing of all occupants of the respective vehicle is inevitable. A differentiation can only be made with regard to the number of fatalities. In this respect, a calculation as the comparison of the sums would be a simple and conceivable criterion.[18]

It would be more difficult if the crash did not lead to the death of all occupants but to injuries of varying severity. This would fall outside the actual scope of our deliberations, namely, the endangerment of at least two human lives and thus the question of the sparing and simultaneous sacrifice

16 Cf. Goodall 2014 or, more recently, Gerdes/Thornton 2015.

17 In this context, the myth of the computer as the embodiment of the ideal of rationality should not be underestimated. Cf.Kuhlmann 1985.

18 The questioning of a summation of deaths is quite justified; however, it is the result of a view that the death (or, in milder terms, the injury) of a person is accepted in order to spare others this death and is thus only a means to an end, in this case the rescue of other persons. This view, however, is one that arises from or corresponds to deontological ethics.

of one or the other human life, since it would then only be about (differing) non-fatal personal injury or about the contrast between fatal and non-fatal injury. However, just such a scenario shows the fundamental problems of utilitarianism most clearly, since questions like the following would now arise: how are the injuries to be weighted? Are they to be summed up in the case of several injured persons? How many non-lethal injuries outweigh one dead person? Can a death be weighed against other things at all? Such scoring may seem odd, but the valuation of injuries and even death is common in the insurance industry (Daniel/Ignatieva/Sherris 2019; Cristea/Mitu 2008). The difference, however, is that here, on the basis of such scoring, concrete decisions are made about for whom an injury or even death should be brought about, and not just whether or not an insurance policy should be effected. The relatively simple operationalization in utilitarianism is thus preceded by serious moral decisions.

The situation becomes even more difficult if not only the consequences of damage are included in the calculation (which then would also point in the direction of risk ethics – but more on this later), but if these are set off against the benefit, as is common in utilitarian approaches in the tradition of Bentham and Mill (Bentham 2013; Mill 2009). While in the case of harm, this could, in the context of personal injury as seen above, be relatively easily calculated from the probability, severity and duration of an injury (up to and including death), and only the scoring poses moral problems, in the case of benefit, the very determination of what is to be considered a benefit is problematic. What is benefit in the context of personal injury anyway? The most obvious, simplest, but perhaps also most trivial answer would be: the absence of harm. But this would define benefit as complementary to harm and vice versa. This is precisely not the case in utilitarianism (Zoglauer 1997: 160ff.). Moreover: a determination of benefit independent of harm would become inevitable at the very latest if the harm were to be the same in the case of both decisions, which would have to be assumed a priori in the scenario of the alternative death of at least one person.

The question of utility in the context of – not only fatal, but especially fatal – personal injury comes down to the question of the individual and thus differential value of people or groups of people. This insight is not only the result of howsoever informed, yet introspective considerations in the sense of monological ethics, but is also substantiated on a broad empirical basis. In particular, the specific studies and surveys of Rahwan et al. have shown that people make valuations about people even, or indeed especially,

in the context of the trolley problem and thus in the case of (fatal) personal injury (Rahwan et al. 2018): for example, people socialized in Western cultures decide against elderly occupants, thus allowing the autonomous vehicle to collide with the vehicle in which elderly people are seated. In the eastern, i.e. Asian, cultural sphere, on the other hand, people decide against younger occupants. Seniority is apparently attributed a higher value there. Should age therefore to be taken into account in the AI's decision? And what about gender? Profession? Or with regard to China: social reputation, determined within the framework of the Social Credit System? Is such culturally dependent scoring justifiable?[19] Or is it, on the contrary, a necessity?[20]

So as not to be misunderstood: the questions raised above regarding the definition of benefit as well as the respective quantification of benefit and harm are not only difficult to answer morally (from a non-utilitarian point of view), but also pragmatically (from a utilitarian point of view). However, once they have been answered or the relevant valuations have been made, the decision can be effected mathematically and mechanically.

For utilitarianism, a decision on the physical integrity of persons – even and especially up to and including a decision on their death – on the basis of a harm-benefit calculation – is completely unproblematic, i.e. legitimate, from a moral point of view (Singer 2013: 207). Accordingly, utilitarianism could, on the one hand, provide the development of accident algorithms with the legitimacy for a machine decision about life and death, while, on the other hand, it generates from the considerations gained through thought experiments a whole scope of criteria to be considered. Whether with the current state of the art it is technically possible to account for every criterion is initially irrelevant: on the one hand, because following an old legal principle[21] what is not possible is not obligatory. Its non-consideration is therefore not a problem in itself. On the other hand, because what is not

19 For a critique of cultural relativism, see Zoglauer 2017: 253ff. and Zoglauer 2021.

20 The problem of whether harm and benefit of an action can be set off against each other at all (Zoglauer 1997: 163f.) will be left aside for the time being (and the reader is referred to the section on risk ethics below), since we are not concerned here with moral problems of utilitarianism but rather, in keeping with the topic, with procedural difficulties, i.e. those concerning operationalization/calculation.

21 This idea appeared in Roman law in various expressions, the variety of which also reflects its significance: "Ultra posse nemo obligatur", "Impossibilium nulla obligatio", "Impotentia excusat legem", "Nemo dat quod non habet", "Nemo potest ad impossibile obligari". Cf. Depenheuer 2014: para. 4.

possible today will be possible tomorrow and the criterion that cannot yet be technically implemented, by the fact that it has been thought of, shows the way for development.[22]

5. Deontological or Duty Based Ethics

While in the utilitarian paradigm a valuation of people and the setting off of their respective worth is no problem, both contradict the deontological prohibition of setting off or instrumentalizing humans in the spirit of Kant. Kant formulates the 'practical imperative' in his 1785 work "Grundlegung der Metaphysik der Sitten" as follows: "Act in such a way that you treat humanity, whether in your person or in the person of any other, never merely as a means, but at all times also as an end" (Kant AA IV: 429/Terence 2009: 45). Twelve years later, this was stated in greater detail in "Metaphysics of Morals":

> But a human being considered as a person, i.e., as a subject of a morally practical reason, is above all price. For as such (homo noumenon) he is not to be valued merely as a means to the ends of other people, or even to his own ends, but as an end in himself. This is to say, he possesses a dignity (an absolute inner worth), by which he exacts respect for himself from all other rational beings in the world, can measure himself against each member of his species, and can value himself on a footing of equality with them. (Kant AA VI: 434f.)

Two facts are decisive in this context: first, a human being (as a person) is "an ends in himself" and must never be misused "merely as a means." What is meant by this is that people may not be instrumentalized to achieve ends other than as ends in themselves, which would be the case in the scenario of sacrificing or harming a human life for the purpose of saving other human lives, even if it be several. The decisive factor is action, the human being as agent, so that, as will become even clearer in the following, the difference between action, or active action, and non-action, or omission,

22 At this stage, or perhaps already at a point beyond it, is probably where the metaverse that Mark Zuckerberg is currently massively advancing with his company Meta (formerly: Facebook) finds itself, in which even Goldman Sachs sees the future ("Goldman Sachs" 2022).

gains the highest relevance in practice. The vehicle AI in our AI thought experiment would, in any case, have to change lanes in order to spare the three occupants, thereby, through the death or injury of one human as a consequence of this action, misusing one human as a means to the end of saving the other three. However, offsets and instrumentalizations of humans are prohibited in the deontological paradigm.

The second important fact that arises from the passage quoted above from "Metaphysics of Morals" is the justification or derivation of the self-purpose of humans: a human is "above all price." This implies, on the one hand, the impossibility of determining the worth of a human being in relation to a commodity or a currency and, on the other hand, or rather as a consequence thereof, that no human being is worth more or less than another (Zoglauer 2017: 62). Thus, the attribution or determination and, consequently, calculation of any kind of benefit of a non-harming of human beings is not permissible. The equality of human beings that goes hand in hand with this absolute human dignity also prohibits the preferential treatment or disadvantaging of human beings, which, as will be seen below, is turned around and used as an argument in risk ethics for a harm-minimizing approach. First, however, let us consider in greater detail the consequences for a decision based on duty-based ethics of the dilemma discussed here.

Kant's prohibition of instrumentalization ultimately boils down to a non-harm principle that Zoglauer describes as follows: "Let nature take its course and do not actively intervene in events if it is foreseeable that innocent people will die as a result of the intervention" (Zoglauer 1997: 73). In the context of accident algorithms, one must not be led astray by the adjective "innocent". Even if the driver of the passing car in our scenario had proceeded to pass despite overtaking being prohibited, he would not be a "guilty" person. The term "innocent" in the aforementioned quote derives from thought experiments previously discussed by Zoglauer, including the scenario of a hijacked airplane with hostages to be used as a terrorist weapon, leading to the question of shooting down or not shooting down the plane, which would not only involve the death of the "guilty" hijackers/terrorists[23], but also of the "innocent" hostages. In our scenario, however, there can be no question of a comparable differentiation. It can be assumed that in the case of road traffic, "innocent" people are always involved. In the case of a dilemma with

23 On the thoroughly problematic concept of terrorist, see Klimczak/Petersen 2015.

necessary (fatal) personal injury, as discussed here for accident algorithms, this would mean that the AI would always be forced to non-action, to passivity. Specifically, the vehicle AI in our example would not be allowed to and would not change lanes, so would have to keep to its lane, resulting in the killing of three persons and not 'only' one person. While in utilitarianism the difference between action and non-action or omission of an action is morally irrelevant – hence action and non-action/omission are completely equal (Wolf 1993: 393) – duty-based ethics, as already touched upon above, rests on a rigorous differentiation, which would always lead to the AI not acting.

This non-action is not to be confused with the AI delegating the decision back to the human driver, which would be tantamount to freeing up the decision. The decision of an AI programmed on the basis of duty-based ethics would already have been made with the decision for non-action. The decision for non-action on the part of the AI is thus accompanied by the impossibility of action on the part of the human, which is much more far-reaching than an analogous decision for a human decision-maker: for a human decision-maker, the instrumentalization prohibition and the non-harm principle merely add up to precisely that: a prohibition or a principle, which can be violated or broken. An 'ought' does not necessarily entail an 'is'; human decision-makers can (and persistently do) disregard prohibitions and can, as such, decide otherwise. For an AI machine, however, an 'ought' goes hand in hand with an 'is'.[24]

This is a complete deprivation of the human being of the right of decision in an existential situation, namely one of life and death. Even if this may seem absurd, this disenfranchisement is actually inherent in the concept of autonomous driving, namely as a positive goal and as the highest achievable level 5 of the previously cited SAE standard for autonomous driving. The reader may remember:

> Level 5: Full Automation – The vehicle is capable of performing all driving functions under all environmental conditions and can operate without humans inside. The human occupants are passengers and need never be involved in driving. A steering wheel is optional in this vehicle. ("Autonomous Driving" 2018)

24 Cf. also and especially from a formal logic perspective Klimczak et al. 2014: 89.

While the SAE does not rule out a steering wheel or other control devices, but also does not regard them to be necessary anymore, this is generally seen as the main feature of Level 5: "Level 5 vehicles do not require human attention – the 'dynamic driving task' is eliminated. Level 5 cars won't even have steering wheels or acceleration/braking pedals" ("The 6 Levels" 2022).

Against the background of a conception of humanity based on autonomy, such a state of affairs may seem untenable, but delegating the decision to the person in the car would not help in such a situation, since, on the one hand, autonomous cars at level 5 should be able to carry out empty journeys, i.e. journeys without human occupants, as well as journeys with persons who do not have the capacity to drive a car, be they children, disabled persons or simply persons without a driving license. On the other hand, as various studies have shown (Merat et al. 2014; Lin 2015: 71f.), even the presence of a person who would be qualified to control the car would not help because, unless they had sufficiently monitored the driving process in advance, they would not be able to survey the situation quickly enough to make a decision that is at least to some extent conscious (albeit intuitive or instinctive). The 'decision' of the human driver, who would be thrown into a life-and-death situation, would thus ultimately amount to a random decision.[25]

Nevertheless, or for this very reason, such non-action of the AI, such necessary non-action, even if it could be easily programmed or executed as a non-action as an accident algorithm, hardly seems acceptable. While in the context of AI decisions or machine decisions the non-acceptability of a rigorous, invariable non-action becomes particularly apparent, because in this way the 'ought' of duty-based ethics becomes a realized 'is', even in the case of the discussion of non-action in the context of human decision-making, invariable non-action has been regarded as unsatisfactory.[26]

6. Hazard Diversion or Hazard Prevention

Accordingly, positions have been taken and proposals made in ethics as to the conditions under which, in the case of harmful non-action, harmful ac-

25 On chance, particularly in the context of machine-produced chance, see below.

26 This discomfort or dissatisfaction is most evident in that alternative proposals are being seriously discussed or suggested in the discourse as solutions, at least some of which will be outlined in the next section.

tion would be permitted as an alternative. The irony is that the decision for harmful action or harmful non-action is ultimately based on utilitarian criteria. For instance, Foot had already argued in her trolley example for acting and thus killing one person while saving the five people who would have been killed in the case of non-action, because more lives can be saved in this way (Foot 1997: 193). The fact that such a damaging, but precisely damage-minimizing, action is permitted is justified by her with the imperative of hazard prevention: it is – in Zoglauer's words – "merely diverting a danger not caused by oneself from a large crowd (five people) to a smaller crowd (one person)" (Zoglauer 2017: 242). Harmful action, if it redirects danger for the purpose of averting danger, is permitted. Anything that is not a mere diversion, on the other hand, would not be permissible:

> So we could not start a flood to stop a fire, even when the fire would kill more than the flood, but we could divert a flood to an area in which fewer people would be drowned. (Foot 1980: 159f.)

In respect to our example and the context accident algorithms, this would mean that a literal 'steering' of the vehicle and the ensuing collision with the one-passenger vehicle in the oncoming lane would be allowed because a hazard, the inevitable collision with one of the two cars, is imminent and the action, the change of lanes, would involve damage minimization.

What, however, would in our context or rather example be the analogous case for the flood that Foot considers not to be covered by the imperative of hazard prevention, namely the one triggered in a damage-minimizing manner in order to contain a fire? In the context of the trolley problem, it is worth referring to the so-called Fat Man thought experiment. This trolley variation, introduced by Thomson, assumes that the five people on the track could be saved by pushing a fat man off a bridge, since he would bring the streetcar to a halt with his corpulence, while, however, dying in the process (1976: 204). In this case, it is scarcely possible to speak of a diversion of danger and thus of hazard prevention: the trolley car, which no longer can be braked, could not otherwise possibly have run over the fat man and thus killed him – this would be different in the case of a flying object hazard. The fat man is a person completely uninvolved in and unrelated to the event.

Although in the context of this thought experiment it is sometimes pointed out that his being pushed off the bridge and thus his death can be seen as a means to the end of saving the five people (e.g. Zoglauer 2017: 78, 80) – which is also true –, this argument also applies to the change of

lanes of the trolley car and the subsequent killing of the one person on the siding. The twist of the imperative of hazard prevention or hazard diversion lies precisely in the fact that it can indeed be used to justify a damaging action that results in a (quantitative) minimization of damage, but in the realization of this the injured party or parties was or were nevertheless turned into a means to an end, namely that minimization of damage. The exclusion of uninvolved or unrelated persons from the requirement to avert danger thus represents a restriction of the permission to act in a harmful manner for the purpose of utilitarian harm minimization.

With regard to the case of accident algorithms being considered here, however, an analogous case can hardly be described: after all, this scenario is literally about steering away from danger. The car with the three occupants approaching in the wrong lane is a participant in road traffic, thus it is not uninvolved.[27] This might also be true in an extended thought experiment for a pedestrian whose damage-minimizing crossing might offer itself as an alternative or third decision option: pedestrians are also road users. And even if the pedestrian were not a road user but were walking on the grass with an appropriate distance to the roadway, it would in this case still be the diverted danger, namely the car, that would run him over and prevent the deaths of the three occupants in the other car. Whether the scenario can still be considered realistic or not is irrelevant: the notion that the fat man could bring a streetcar to a halt with his corpulence is also highly improbable. The point is merely to show that all those potentially injured by an AI's driving decisions are involved or affected when it comes to danger prevention.

If we focus on the difference between participation and non-participation in the Fat Man thought experiment, no restriction on the prohibition of danger prevention and thus also no restriction on the permission to act in a harmful way can be deduced for the original trolley scenario. However, the Fat Man thought experiment presents another difference to the classical trolley case: it is not the trolley driver who performs the action, but a third party outside the object of danger, in this case outside the trolley. Thus, one could say that not only the injured party is a bystander, but also the injuring

27 Unless one defines bystanders by the fact that they do not come to harm in the case of the choice of omission, which would be an extremely trivial definition and would also lead to one never being able carry out a diversion of danger, which would make no sense in the context of the discussion of exceptions of danger diversions or would render them obsolete because no danger diversion would be permitted at all.

party. The injurer in the Fat Man case is not directly part of the danger, but nevertheless decides. Against this backdrop, the question of the status of the decision-making AI in the autonomous vehicle is almost unavoidable. Can an entity that is not a living being as a machine and can therefore suffer material damage but not personal injury still be said to be involved or affected? [28]

If the (direct) involvement of the injuring party in a hazard were to be regarded as a necessary prerequisite for the hazard prevention imperative and if the AI in the autonomous vehicle could not be assumed to be a direct participant, then the concept of hazard prevention and thus the permission to actively inflict harm for a harm-minimizing purpose would be ruled out for accident algorithms. What is at question is whether the involvement of the injuring party is actually a prerequisite: Foot's flood example, after all, assumes a person diverting the flood whose participation or affectedness is not further specified. And regardless of how exactly participation or affectedness of the damaging party is defined, the definitions will always be those for human decision-makers and not for machine decision-makers, whose being involved or affected is of an essential, i.e. metaphysical or ontological nature. In other words: either the imperative of hazard prevention leads to the fact that AI, being AI, will have to decide for non-action, as in the case of strict duty-based ethics, or to the fact that decisions will be made according to a harm-minimizing approach.[29]

---

28 In addition, it can be assumed that an AI in an autonomous vehicle will not be a local AI, but a decentralized cloud AI that is not on site at all, which means that participation is also not spatially given.

29 Regardless of whether the actual utilitarian approach is rejected due to the impossibility of offsetting benefits and harms or the problem of quantifying and defining benefits (see above) since this necessarily presupposes an individual-specific valuation of people, a harm-minimizing approach with risk ethics seems to be gaining ground, especially in the context of autonomous driving. The circle around Julian Nida-Rümelin (Hevelke/Nida-Rümelin 2015) does not, at any rate, see this as a violation of Kant's prohibition of instrumentalization. Other modifying concepts such as that of voluntariness, which have been designed and discussed for medical scenarios (Zoglauer 2017: 108ff.), do not need to be discussed here, since either no one in road traffic would explicitly accept consequences of harm voluntarily, or everyone in road traffic would have to accept consequences of harm voluntarily by implication of their participation in road traffic.

7. The Non-binding Nature of Ethics

Recapitulating what has been presented so far, it can be said: a utilitarian benefit-harm calculation is not justifiable on the basis of deontological ethics; the non-harm principle of deontological ethics, in turn, violates the utilitarian approach of thinking about a decision from the point of view of consequences. Thus, both ethical paradigms contradict each other – and not only that: both utilitarianism and duty-based ethics reach moral limits when viewed unbiasedly. Since ethics is not a natural science in which, due to the object area, a correspondence-theoretical approach between scientific statements and reality[30] is at least to some extent given, there is also in principle no reason in case of contradictory models of thought to ascribe a higher quality to the one than to the other.[31] And even models of truth and validity based on consensus and coherence (Zoglauer 2007: 250ff.) do not lead to success with regard to deciding in favor of one or the other ethics model.

To put it more succinctly: there is no compelling, necessary reason why the deontological notion of a prohibition of instrumentalization should be accepted without reservation. And the presented concepts of hazard prevention or diversion or those of risk ethics have shown (in an exemplary way) that possibilities for a restriction or a combination of deontological and utilitarian approaches are at least being sought or seriously entertained and thought through in the deontological paradigm. Regardless of whether a rigorous utilitarianism, a rigorous duty-based ethics, or a combination of both approaches is considered valid and thus action-guiding for specific decision-making and, in this context, for technical development and innovation, these are decisions, positings, none of which are obligatory.

Against the background of discursive theories of truth and decision, there have always been proposals for a procedural model in ethics – e.g. Rawls' (1951) "competent moral judges" who should make binding moral judgments or decisions – but it has never advanced as far as an institutionalization of ethical authorities. This is hardly surprising, since ethics is

30 Even if one were to accept the statistically recorded and calculated moral perception of society as just that very empirical basis, this is, as touched on in the text and footnotes above, at least in part culturally relative, both synchronously between cultures and diachronically within a culture.

31 On this and the following aspect, see Zoglauer 1997: 229–274.

a science, and even though certain schools of thought dominate in every science and their dominance is not only rationally justified but also strongly dependent on social factors (Kuhn 1962), science sees itself as free and open, which runs counter to explicit institutionalization.

8. Interdependencies between Ethics, Law and Society I

The lack of institutionalization in ethics and an associated procedural model for moral decisions/judgments is, however, also logical for another reason: with law, specifically law-making and jurisprudence, institutions already exist which, on the one hand, make judgments or decisions concerning morality and, on the other hand, are most closely intermingled or linked with ethical paradigms and discourses in ethics. Legislation – at least in functioning democracies – is based either directly or indirectly, at best ex ante, but at the least ex post, on a societal consensus as part of a societal discourse (Kunz/Mona 2015: 6ff.). Admittedly, this discourse might only in the rarest cases obey the rules of an ideal discourse as developed by Habermas (1981a, 1981b, 1984), Alexy (1994, 1995, 2004) et al. in the context of discourse ethics, precisely because in pluralistic industrial and post-industrial representative democracies not all discourse participants can – neither qualitatively nor quantitatively – have their say equally, be heard equally, have equal opportunity to access information and also actually participate in the discourse.[32] Moreover, topics and opinions that are considered taboo are virtually precluded from discourse, and opinions and views in general are even viewed with prejudice from the outset. Although, therefore, even in democracies there are no discourses free of domination in the sense of Habermas, in democratic systems with a reasonably functioning interplay of executive and legislative decision-making on the basis of the determination by voters of the majority composition of the legislature,[33] the will of society is reflected in legislation, i.e. in lawmaking.

This linkage is, therefore, of essential significance, since otherwise, due to the largely alternative-free, legally positivist functioning of legislation, it

32 On real discourses or practical discourses as distinct from ideal discourses (especially in the context of law), cf. Alexy 1995: 123ff.

33 Which can be provided for to a certain extent, in purely functional terms, by a sufficient alternation of government and opposition.

would be disconnected from the societal discourse. Extensive, positivistic lawmaking is, at least practically, without alternative because otherwise, on the one hand, legal decisions in the absence of legislation would be arbitrary, since they would depend on the individual sense of justice of the decision-maker (Kriele 1991: 135) and, on the other hand, legislation according to natural law would be constantly susceptible to the danger of the naturalistic fallacy (Zoglauer 1997: 204), i.e. the deduction of 'ought' from 'is'. Thus, however – i.e. due to the linkage of essentially positivist lawmaking to the societal discourse in democratic systems – the expert discourse on ethics can find its way into legislation, on the one hand, because it is itself connected to the broad societal discourse in a complex, diffuse interplay, and, on the other hand, because in the case of an indirect democracy it can indirectly influence lawmaking in the form of expert commissions and the advisory systems of political decision-makers.[34] Therefore, even if no ethical paradigm has any more claim to validity than another, one paradigm can and will dominate the ethical discourse and then most likely prevail in the societal discourse, thus ultimately forming the basis of institutionalized lawmaking.[35/36]

However, it is not only lawmaking that is institutionalized and standardized, but also a second, quite inseparable component, namely case law: statutory law, i.e. laws, not only can but also do necessarily contradict each other; this is, on the one hand, because different dominant ethical paradigms that existed at different times over the course of perpetual lawmaking formed the basis of the legislation, and, on the other hand, because the contradictory nature of the laws is even intended, since, by virtue of institutionalized case law, there is the possibility to weigh up and decide on a case-specific basis. The fundamental rights addressed in the case of accident algorithms, in particular, contradict or clash with each other to the greatest possible extent.

34 In the context of autonomous driving, see in particular the ethics committees of the German Federal Ministry of Transport, which, however, will not be considered in detail here, since their statements are not only non-binding but in the main also very vague; this must almost inevitably be the case, not only because the ethical discourse is pluralistic, but also, as described above, no school of thought can lay claim to (absolute) truth and will, therefore, in the normal case, i.e. with a heterogeneous composition of committee members, not be able to establish itself.

35 This will become clear in the next section when we look at legal issues.

36 In this respect, I disagree with the positivist understanding of law and morality as two independent norm spheres (cf. Kelsen 1979: 102 and, as overview, Zoglauer 1997: 133).

Regardless of the legal discourse on whether laws actually contradict each other in formal/formal-logical terms, only seem to contradict each other or do not contradict each other at all,[37] law has developed a number of derogation principles, such as the lex superior derogat legi inferiori rule, according to which a higher level rule abrogates a lower one, the lex specialis derogat legi generali rule, according to which a special rule takes precedence over a general one, or the lex posterior derogat legi priori, according to which newer rules override older ones.[38]

However, these derogation principles do not apply or help in cases where the conflicting rules have the same level of relevance or generality or were enacted at the same time. All this is not only the case with the aforementioned fundamental rights, but as prima facie norms they are even designed to be so: they are characterized precisely by the fact that their validity is systemically limited by other norms with which they compete, i.e. which contradict them (Alexy 1994: 87ff.). If the facts of a case indicate the application of two competing prima facie norms resulting in a conflict of principles, judicial balancing must come into play. A court must decide which norm takes precedence in the case in question. Kelsen allows this primacy of judicial decision to apply not only in the case of prima facie norms, but also when a judge's decision runs counter to established derogation principles (Kelsen 1979: 175ff.). The problem of a possible arbitrariness of judicial decisions associated with this is mitigated, at least to some extent, by the existence of several judicial instances and the possibility of taking legal action against judicial decisions.

Even if the judgment of the respective highest court must be accepted, regardless of whether it is seen to be just or unjust, and this judgment, moreover, will have, as a decision of the highest court, model character for future judgments at a lower level and thus become law itself, so to speak,[39] one cannot go so far as to claim that the whole of law "depends solely on the volition of judicial organs" (Zoglauer 1997: 136). At least in the medium and long term, in the event that a supreme court decision violates society's sense of justice, this sense of justice will assert itself through the influence

37 For a detailed sketch of this discourse from the perspective of the formal logician, see Zoglauer 1997: 125–146.

38 On derogation in detail, cf. Wiederin 2011: 104ff.

39 See Schick 2011 for a detailed discussion of the topics of interpretation and further development of the law.

of societal discourse[40] on the composition and decisions of the legislative branch as well as the filling of vacancies in the respective highest court or will at least find expression in some other milder form.[41]

9. The Judgment on the Aviation Security Act

As far as is apparent, there is no treatise dealing with the trolley problem in the context of autonomous driving that does not also extensively refer to or at least mention the ruling of the German Federal Constitutional Court (BVerfG) of February 15, 2006 on the Aviation Security Act (LuftSiG). The background to the LuftSiG was the attacks of September 11, 2001, in the United States and the debate that began in Germany in the aftermath as to how preventive action could be taken in a similar scenario (Feldle 2018: 146): in the event that terrorists take control of an airplane with uninvolved passengers and want to using it as a weapon by crashing it into a built-up area, would it be legal to shoot down the plane and thus sacrifice not only the lives of the terrorists but also those of the innocent passengers in order to save other human lives? This is, in any case, what the legislature wanted to create the legal basis for with the LuftSiG.

The fact that the scenario with the hijacked, presumably unrescuable passengers that is behind the Aviation Security Act is nevertheless applicable to the trolley problems of autonomous driving is thanks to the argumentation of the BVerfG. The court explicitly does not share the view that the lives of the passengers in the hijacked aircraft are already forfeited and speaks in this regard of the "uncertainties as regards the factual situation"[42]: either the passengers could take control of the aircraft and render the hijackers harmless, or the hijackers themselves could change their minds. Though unlikely, neither of these is out of the question. In the opinion of the BVerfG, there can be no question of "these people's lives being 'lost anyway already'".[43] This means, however, that the chances of rescue are no longer absolute, but only

40 Cf. again Kunz/Mona 2015: 6ff.

41 On the multifaceted interdependencies between the Federal Constitutional Court and the political system of the Federal Republic, cf. the numerous articles in the comprehensive handbook Van Ooyen/Möllers 2015.

42 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 133.

43 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 133.

gradually asymmetrical in relation to the probability of a rescue occurring. The airplane scenario is thus similar to the trolley scenario in that here a group of people is no longer irretrievably lost, so it is not merely a matter of saving or not saving people, in this case another group of people (since the highjacked people are lost anyway), but of sacrificing people in favor of saving other people. That these two scenarios are not identical but merely approximate each other is due to the fact that, if one takes the argumentation of the BVerfG to its logical end, no one has to come to any harm in the airplane scenario, so there could also be no fatalities at all and it is therefore precisely not a dilemma – quite in contrast to the trolley dilemma of autonomous driving.

This difference, which continues to exist but is not discussed in the literature, does not seem to be a problem because the court also deals with the scenario in which the lives of the aircraft passengers are already forfeited, which is ultimately only hypothetical for the court but is foundational for the legislature. The BVerfG argues with regard to "the assessment that the persons who are on board a plane that is intended to be used against other people's lives within the meaning of § 14.3 of the Aviation Security Act are doomed anyway"[44] that these people would also enjoy the "same constitutional protection". This would exist "regardless of the duration of the physical existence of the individual human being."[45] A diverging opinion of the state, according to the court, would render these people as mere objects, not only in the service of the kidnappers, but precisely in service of the state:

> By their killing being used as a means to save others, they are treated as objects and at the same time deprived of their rights; with their lives being disposed of unilaterally by the state, the persons on board the aircraft, who, as victims, are themselves in need of protection, are denied the value which is due to a human being for his or her own sake.[46]

Even the motive of saving other people by sacrificing the hijacked passengers does not change this: "Finally, § 14.3 of the Aviation Security Act also cannot be justified by invoking the state's duty to protect those against whose lives the aircraft that is abused as a weapon for a crime within the meaning of §

44 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 132.

45 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 132.

46 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 124.

14.3 of the Aviation Security Act is intended to be used."[47] Or: "The fact that this procedure is intended to serve to protect and preserve other people's lives does not alter this."[48]

This reasoning of the BVerfG is interpreted such that in the case of "irretrievably lost life […] the person concerned is left with the least possible remaining time" and if "even the sacrifice of this miniscule remainder of life cannot be justified, […] then this applies a fortiori to the life of an old person whose exact date of death is impossible to predetermine" (Feldle 2018: 150f.). And since the qualitative differentiation of the choice of victim according to age can only be about the remaining time of life (Feldle 2018: 150f.), all qualitative valuations of human beings are thus prohibited (Feldle 2018: 152).

In its ruling, however, the Federal Constitutional Court not only opposed qualitative offsetting, but also, if only implicitly, quantitative offsetting, since it explicitly reproduces the Bundestag's view that "having a relatively smaller number of people killed by the armed forces in order to avoid an even higher number of deaths"[49] does not violate Article 2 (2) sentence 1 of the Basic Law and then the Court goes on to describe in detail[50] that the Aviation Security Act violates the Basic Law precisely because of Article 2 (2) sentence 1 of the Basic Law. The statements that it is 'only' about the prevention of "the occurrence of especially grave accidents within the meaning of Article 35.2 sentences 2 and 3 of the Basic Law"[51] and not "protecting the legally constituted body politic from attacks which are aimed at its breakdown and destruction"[52] can also be read along these same lines: the decisive factor would be a particular quality of consequence of the act, in this case the breakdown and destruction of the body politic, and not a quantity of consequence of some kind, measured in terms of human lives.[53]

In all conceivable cases, the BVerfG thus – as can be seen from the passages already quoted – explicitly cites Kant's prohibition of instrumentalization, on the one hand, and cites fundamental rights on the other:

47 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 137.
48 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 139.
49 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 48.
50 BVerfG, Judgment of February 15, 2006 – 1 BvR 357/05, para. 75ff.
51 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 136.
52 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 135.
53 Even though this view is less justified than posited, cf. Feldle 2018: 153.

specifically, it states right at the beginning of the reasoning for the judgment[54] and repeatedly in what follows that a shooting down of the aircraft is incompatible with the right to life guaranteed in Art. 2 Abs. 2 S. 1 GG ("Everyone has the right to life and physical integrity.") in connection with the guarantee of human dignity from Article 1 paragraph 1 GG ("Human dignity shall be inviolable. To respect and protect it shall be the duty of all state authority."). Accordingly, the BVerfG ruled that Section 14 (3) of the LuftSiG was "The regulation is completely unconstitutional and consequently, it is void pursuant to § 95.3 sentence 1 of the Federal Constitutional Court Act"[55].

10. Interdependencies between Ethics, Law and Society II

The argumentation as well as the outcome of the judgment was partly met in the jurisprudential discourse with plain and strident criticism (Gropp 2006; Hillgruber 2007; Sinn 2004), not least also from a pragmatic point of view, because the state is thus forced to take no action and, in this respect, the concept of strict duty-based ethics has thereby prevailed. Nevertheless, the fact that the relevant part of the Aviation Security Act was declared unconstitutional and thus null and void and that the legislature enacted a new version of §14 LuftSiG that no longer allows for the preventive shooting down of a passenger plane, would suggest that the whole process contradicts the remarks above on the societal influence on jurisprudence or its interplay with societal discourse: not only do decisions of the highest courts have to be accepted as case law or de facto law – decisions of the Constitutional Court, in this case the Federal Constitutional Court, can even nullify laws that have been formally correctly passed by the legislature. Since the basis of the judgment was formed by Article 1 and Article 2 of the Basic Law, i.e. those very articles that are unchangeable under Article 79 (3) of the Basic Law, no constitutional amendment on the part of the legislature could have swayed the Federal Constitutional Court to change the basis of the judgment. To avoid any misunderstanding: my point is not that this would be desirable, but rather that jurisdiction in this regard is removed from the legislature and thus from that instance to which society, in the form of the electorate, has the most direct access.

54 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 81.

55 BVerfG, judgment of February 15, 2006 – 1 BvR 357/05, para. 155.

Nevertheless, it must be pointed out that even if the judgments of the Federal Constitutional Court are less dependent on the will of the people or, in the terminology used here, more decoupled from the respective societal discourse, the judges nevertheless argued with Kant and thus with a philosopher who can be considered the most influential in Germany's ethical discourse. In other legal systems this is precisely not the case. In Anglo-Saxon legal systems, for example, the utilitarian view is certainly advocated in decisions about life and death (LaFave 2010: 557; Ormerod 2011: 371), which shows that decisions to this effect are certainly influenced by ethical and thus indeed elitist, but nonetheless societal, discourses.

Moreover, not only supreme court opinions but also those of the Federal Constitutional Court with regard to fundamental rights can change, as can be clearly seen in the *quantitative* offsetting of human lives. Even if the view that human lives must not be weighed against each other is virtually unquestioned in German legal opinion today and is based on the fundamental rights of the Basic Law, it only became established in the 1950s as part of the negotiations surrounding the euthanasia murders by the Nazi regime (Feldle 2018: 155), hence at a time when the Basic Law already existed. This, in turn, shows that an interpretation of fundamental rights as standing in contradiction to quantitative offsetting is not compelling. Lastly, but not least importantly, although fundamental rights are protected from amendment by virtue of the aforementioned 'eternity clause,' i.e. Article 79(3) of the Basic Law, this 'eternity' is tied to the legal force of the Basic Law. According to Article 146 of the Basic Law, the German people may still freely decide on a (new) constitution.[56]

In summary, even in the area of fundamental rights, the influence of societal discourse can be detected, albeit very directly and only in the long term, such that although legislation and case law are of primary importance for the development of technology, societal discourse in general and ethical discourse in particular open up the space for thought and possibility and

56 Interestingly, after German reunification, the path via Article 146 of the Basic Law was not chosen because, among other things (!), a weakening of fundamental rights was feared at the time.

provide stimuli, which are then reflected in legislation and case law in the medium and long term.[57]

References

Abele, Thomas (ed.) (2019). *Case Studies in Technology & Innovation Management*, Wiesbaden.

Albers, Sönke, and Oliver Gassmann (eds.) (2005). *Handbuch Technologie- und Innovationsmanagement*, Wiesbaden.

Alexy, Robert (1994). *Theorie der Grundrechte*, Frankfurt/Main.

Alexy, Robert (1995) *Recht, Vernunft, Diskurs. Studien zur Rechtsphilosophie*, Frankfurt/Main.

Alexy, Robert (2004). *Theorie der juristischen Argumentation*, Frankfurt/Main.

"Autonomous Driving – Hands on the Wheel or No Wheel at All." April 11, 2018. URL: https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all/#gs.vctpcs.

"Autonomes Fahren: Das bedeuten Level 0 bis 5." June 15, 2021. URL: https://t3n.de/news/autonomes-fahren-assistenz-fahrautomatisierung-level-stufe-1332889/.

Bentham, Jeremy (2013). *An Introduction to the Principles of Morality and Legislation*, Saldenberg.

Berkemer, Rainer, and Markus Grottke (2023). "Learning Algorithms – what is artificial intelligence really capable of?" In: *Artificial Intelligence – Limits and Prospects*. Ed. by Peter Klimczak and Christer Petersen, Bielefeld 2023.

Chun, Wendy Hui Kyong (2018). "Queerying Homophily. Patterns of Network Analysis." In: *Journal of Media Studies* 18, pp. 131–148.

Cristea, Mirela, and Narcis Eduard Mitu (2008). "Experience Studies on Determining Life Premium Insurance Ratings: Practical Approaches." In: *Eurasian Journal of Business and Economics* 1.1, pp. 61–82.

Daniel, Alai, Katja Ignatieva, and Michael Sherris (2019). "The Investigation of a Forward-Rate Mortality Framework." In: *Risks* 7.2, (https://doi.org/10.3390/risks7020061).

57 Conversely, it can, therefore, be said that trends in societal and ethical discourse on the part of technical development should certainly be perceived, because this allows us to anticipate what future regulations might look like.

Datta, Amit (2019). "Die Haftung bei Verträgen über digitale Inhalte." In: *Recht als Infrastruktur für Innovation*. Ed. by Lena Maute and Mark-Oliver Mackenrodt (eds.), Baden-Baden, pp. 155–178.

Depenheuer, Otto (2014). "§ 269 Vorbehalt des Möglichen." In: *Handbuch des Staatsrechts der Bundesrepublik Deutschland, vol. XII: Normativität und Schutz der Verfassung*. Ed. by Josef Isensee and Paul Kirchhof, pp. 557–590.

Dilich, Michael A., Dror Kopernik, and John Goebelbecker (2002). "Evaluating Driver Response to a Sudden Emergency: Issues of Expectancy, Emotional Arousal and Uncertainty." In: *SAE Transactions* 111, pp. 238–248.

Engisch, Karl (1930). *Untersuchungen über Vorsatz und Fahrlässigkeit im Strafrecht*, Berlin.

Färber, Berthold (2015). Communication Problems between Autonomous Vehicles and Human Drivers. In: *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte*. Ed. by Markus Maurer et al., Wiesbaden, pp. 127–150.

Feldle, Jochen (2018). *Notstandsalgorithmen. Dilemmata im automatisierten Straßenverkehr*. Baden-Baden.

Foot, Philippa (1980). "Killing, Letting Die, and Euthanasia: A Reply to Holly Smith Goldman." In: *Analysis* 41.3, pp. 159–160.

Foot, Philippa (1967). "The Problem of Abortion and the Doctrine of the Double Effect." In: *Oxford Review* 5, pp. 5–15.

Gerdes, J. Christian, and Sarah M. Thornton (2015). "Implementable Ethics for Autonomous Vehicles." In: *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte*. Ed. by Markus Maurer et al., Wiesbaden, pp. 88–102.

"Goldman Sachs sieht Zukunft im Metaversum und Web3." March 29, 2022. URL: https://cvj.ch/invest/goldman-sachs-sieht-zukunft-im-metaversum-und-web3/.

Goodall, Noah J. (2014). "Ethical decision making during automated vehicle crashes." In: *Journal of the Transportation Research Board* 2424, pp. 58–65.

Gropp, Walter (2006). "The Radar Technician Case – A Defensive Emergency Triggered by Humans? Ein Nachruf auf § 14 III Luftsicherheitsgesetz." In: *GA [= Goltdammer's Archiv für Strafrecht]* 153.5, pp. 284–288.

Habermas, Jürgen (1981a). *Theorie des kommunikativen Handelns I: Handlungsrationalität und gesellschaftliche Rationalisierung*, Frankfurt/Main.

Habermas, Jürgen (1981b). *Theorie des kommunikativen Handelns II: Zur Kritik der funktionalistischen Vernunft*, Frankfurt/Main.

Habermas, Jürgen (1984). *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*, Frankfurt/Main.

Hevelke, Alexander, and Julian Nida-Rümelin (2015). "Self-driving cars and trolley problems: On Offsetting Human Lives in the Case of Inevitable Accidents." In: *Yearbook of Science and Ethics* 19, pp. 5–23.

Hillgruber, Christian (2007). "Der Staat des Grundgesetzes – nur 'bedingt abwehrbereit'? Plädoyer für eine wehrhafte Verfassungsinterpretation." In: *JZ – JuristenZeitung* 62.5, pp. 209–218.

Hoffmann-Riem, Wolfgang (2016). *Innovation und Recht – Recht und Innovation. Recht im Ensemble seiner Kontexte*, Tübingen.

Hörnle, Tatjana, and Wolfgang Wohlers (2018). "The Trolley Problem Reloaded, How are autonomous vehicles to be programmed for life-versus-life dilemmas?" In: *GA [=Goltdammer's Archive for Criminal Law]* 165.1, pp. 12–34.

Kalberg, Stephen (1980). "Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History." In: *American Journal of Sociology* 85.5, pp. 1145–1179.

Kant, Immanuel (1900ff). *Gesammelte Schriften*. Ed. by the Prussian Academy of Sciences. Berlin. URL: https://korpora.zim.uni-duisburg-essen.de/kant/verzeichnisse-gesamt.html

Kelsen, Hans (1979). *Allgemeine Theorie der Normen*, Vienna.

Kleinschmidt, Sebastian, and Bernardo Wagner (2020). "Technik autonomer Fahrzeuge." In: *Autonomes Fahren. Rechtsprobleme, Rechtsfolgen, technische Grundlagen*. Ed. by Bernd Opperman and Jutta Stender-Vorwachs, Munich, pp. 7–38.

Klimczak, Peter, Isabel Kusche, Constanze Tschöpe, and Matthias Wolff (2019). "Menschliche und maschinelle Entscheidungsrationalität – Zur Kontrolle und Akzeptanz Künstlicher Intelligenz." In: *Journal of Media Studies* 21, pp. 39–45.

Klimczak, Peter, and Christer Petersen (2015). "Amok. Framing Discourses on Political Violence by the Means of Symbolic Logic." In: *Framing Excessive Violence*. Ed. by Marco Gester, Steffen Krämer, and Daniel Ziegler, Basingstoke, pp. 160–175.

Kriele, Martin (1991). "Recht und Macht." In: *Einführung in das Recht*. Ed. by Dieter Grimm, Heidelberg, pp. 129–171.

Kühling, Jürgen (2009). "Innovationsschützende Zugangsregulierung in der Informationswirtschaft." In: *Innovationsfördernde Regulierung*. Ed. by Martin Eifert and Wolfgang Hoffmann-Riem, Berlin, pp. 47–70.

Kuhlmann, Stefan (1985). "Computer als Mythos." In: *Technology and Society* 3, pp. 91–106.

Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*, Chicago.

Kunz, Karl-Ludwig, and Martino Mona (2015). *Rechtsphilosophie, Rechtstheorie, Rechtssoziologie. Eine Einführung in die theoretischen Grundlagen der Rechtswissenschaft*, Bern.

LaFave, Wayne (2010). *Criminal Law*, St. Paul.

Lemke, Thomas (2000). "Neoliberalism, State and Self-Technologies. A Critical Overview of 'Governmentality Studies.'" In: *Politische Vierteljahresschrift* 41.1, pp. 31–47.

Lin, Patrick (2015). "Why Ethics Matters for Autonomous Cars." In: *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte*. Ed. by Markus Maurer et al., Wiesbaden, pp. 69–85.

Merat, Natasha, Hamish Jamson, Frank Lai, and Oliver Carsten (2014). "Human Factors of Highly Automated Driving: Results from the EASY and CityMobil Projects." In: *Road Vehicle Automation*. Ed. by Gereon Meyer and Sven Beiker, Cham, pp. 113–125.

Meyer, John W., and Ronald L. Jepperson (2000). "The 'Actors' of Modern Society. The Cultural Construction of Social Agency." In: *Sociological Theory* 18.1, pp. 100–120.

Mill, John Stuart (2009). *Utilitarianism*, Hamburg.

Möhrle, Martin G., and Ralf Isenmann (eds.) (2017). *Technologie-Roadmapping. Future Strategies for Technology Companies*, Wiesbaden.

Ormerod, David (ed.) (2011). *Smith and Hogan's Criminal Law*, New York.

Rahwan, Iyad et al. (2018). "The Moral Machine experiment." In: *Nature* 563, pp. 59–64.

Rawls, John (1951). "Outline of a Decision Procedure for Ethics." In: *The Philosophical Review* 60, pp. 177–197.

"Rolling out driverless cars is 'extraordinary grind', says Waymo boss." January 4, 2021. URL: https://www.ft.com/content/6b1b11ea-b50b-4dd5-802d-475c9731e89a.

Roßnagel, Alexander (2009). "'Technikneutrale' Regulierung: Möglichkeiten und Grenzen." In: *Innovationsfördernde Regulierung*. Ed. by Martin Eifert and Wolfgang Hofmann-Riem, Berlin, pp. 323–338.

Schäffner, Vanessa (2020). "When ethics becomes a program: A Risk Ethical Analysis of Moral Dilemmas of Autonomous Driving." In: *Journal of Ethics and Moral Philosophy* 3, pp. 27–49.

Schick, Robert (2011). "Auslegung und Rechtsfortbildung." In: *Rechtstheorie. Rechtsbegriff – Dynamik – Auslegung*. Ed. by Stefan Griller and Heinz Peter Rill, Vienna, pp. 209–222.

Schuster, Frank Peter (2017). "Notstandsalgorithmen beim autonomen Fahren." In: *RAW [=Recht Automobil Wirtschaft]*, pp. 13–18.

Servatius, Michael (2019). "Das Verhältnis von Urheber und Verlag – Wohin steuert die Verlagsbeteiligung?" In: *Recht als Infrastruktur für Innovation*. Ed. by Lena Mauta and Mark-Oliver Mackenrodt, Baden-Banden, pp. 201–221.

Singer, Peter (2013). *Praktische Ethik*, Stuttgart.

Sinn, Arndt (2004). "Tötung Unschuldiger auf Grund § 14 III Luftsicherheitsgesetz – rechtmäßig?" In: *Neue Zeitschrift für Strafrecht* 24.11, pp. 585–593.

"The 6 Levels of Vehicle Autonomy Explained". URL: https://www.synopsys.com/automotive/autonomous-driving-levels.html.

Thomson, Judith Jarvis (1976). "Killing, Letting Die, and the Trolley Problem." In: *The Monist* 59, pp. 204–217.

Van Ooyen, Robert, and Martin Möllers (2015). *Handbuch Bundesverfassungsgericht im politischen System*, Wiesbaden.

Wagner, Helmut (1997). "Rechtsunsicherheit und Wirtschaftswachstum." In: *Ordnungskonforme Wirtschaftspolitik in der Marktwirtschaft*. Ed. by Sylke Behrends, Berlin, pp. 227–254.

Wiederin, Ewald (2011). "Die Stufenbaulehre Adolf Julius Merkls." In: *Rechtstheorie. Rechtsbegriff – Dynamik – Auslegung*. Ed. by Stefan Griller and Heinz Peter Rill. , Vienna, pp. 81–134.

Wolf, Jean-Claude (1993). "Active and Passive Euthanasia." In: *Archiv für Rechts- und Sozialphilosophie* 79, pp. 393–415.

Wördenweber, Burkard, Marco Eggert, Größer, André, and Wiro Wickord (2020). *Technologie- und Innovationsmanagement im Unternehmen*, Wiesbaden.

Zoglauer, Thomas: *Normenkonflikte – zur Logik und Rationalität ethischen Argumentierens*. Stuttgart – Bad Cannstatt 1997.

Zoglauer, Thomas (2017). *Ethical Conflicts between Life and Death. On hijacked airplanes and self-driving cars*. Hannover.

Zoglauer, Thomas (2021). "Wahrheitsrelativismus, Wissenschaftsskeptizismus und die politischen Folgen." In: *Wahrheit und Fake im postfaktischdigitalen Zeitalter. Distinctions in the Humanities and IT Sciences*. Ed. by Peter Klimczak and Thomas Zoglauer, Wiesbaden, pp. 21–26.