Antje von Ungern-Sternberg

9. Autonomous Driving: Regulatory Challenges Raised by Artificial Decision-Making and Tragic Choices

Erstveröffentlichung in: Woodrow Barfield and Ugo Pagallo (eds.), Research Handbook on the Law of Artificial Intelligence, 2017, S. 251-278

I. INTRODUCTION

Autonomous cars are among the most fascinating and visible examples of how artificial intelligence will change our daily life. Very soon, autonomous cars will be able to drive safely on public roads without control of a human driver. The technology—allowing the car's computer system to collect data from sensors, to interact with other vehicles, to analyze data and to control the vehicle's function—has already been developed. Currently, self-driving cars are still being tested, but companies like Ford, Google, Mercedes-Benz, Tesla, and Uber, have announced an intention to sell fully autonomous cars and trucks by 2021.[1] It is unclear how fast the new technology will spread. Some expect a very quick disruption in transportation,[2] others forecast an evolutionary deployment scenario, which means that functions of driving assistance, e.g. lane keeping assistance or emergency braking assistance, and of partial automation, e.g. automated parking or automated highway cruising, are gradually integrated into traditional cars until these are eventually replaced by fully autonomous cars.[3]

One can reasonably expect that autonomous cars will greatly enhance road traffic safety, mobility and convenience. Safety will improve as human errors—currently accountable for over 90 percent of all accidents—are avoided.[4] Self-driving cars will comply with road traffic rules unlike human drivers who tend to ignore many rules, and they are in many respects better than human drivers in collecting data, namely by camera, laser (LIDAR), radar, ultrasonic sensors, GPS and by wireless interaction with other cars (V2V) and infrastructure (V2I), in analyzing the data and in reacting quickly in dangerous situations. Autonomous cars may

[1] Susan Hassler, "2017: The Year of Self-Driving Cars and Trucks" December 30, 2016, *IEEE Spectrum* http://spectrum.ieee.org/transportation/advanced-cars/2017-the-year-of-selfdriving-cars-and-trucks.

[2] James Arbib/Tony Seba, *Rethinking Transportation 2020-2030. A RethinkX Sector Disruption Report*, May 2017, www.rethinkx.com/transportation (based on the assumption that self-driving vehicles will boost transport as a service).

[3] Sven Beiker, "Deployment Scenarios for Vehicles with Higher-Order Automation" in Markus Maurer, J. Christian Gerdes, Barbara Lenz and Hermann Winner, *Autonomous Driving* (SpringerOpen 2016) 193, 195.

[4] Thomas Winkle, "Safety Benefits of Automated Vehicles: Extended Findings from Accident Research for Development, Validation and Testing" in *Autonomous Driving, supra* note 3, 335, 354.

transport passengers who are unable to drive a car, for example elderly people, children, or people with disabilities, thereby increasing individual mobility. Finally, former car-drivers will be able to spend the time of their ride more conveniently with other activities like work, hobbies, or recreation.

But autonomous driving might also have negative consequences. It is very likely, for example, that human driving will be outlawed altogether at some point in order to eliminate the risk caused by the human factor. This would of course bar fervent car drivers from experiencing the joy of driving.[5] More importantly, the impact on the environment is to be determined. Autonomous driving could have the positive effect to save energy if smart traffic and passenger management avoided congestion and reduced overall road traffic. However, if all the people who are currently unfit to drive a car will enjoy riding autonomous cars in the future, this might lead to a significant increase in traffic and negative consequences for the environment, to name two concerns.[6]

Turning to legal aspects, road traffic law is a very densely regulated area of law which protects important goods like road safety and traffic fluidity. Traditionally, it is the human driver who must follow the rules of road traffic law. In an autonomous car, it is no longer a human, but an algorithm, i.e. a step-by-step procedure for solving a problem used by a computer,[7] which governs the car's behavior. Shifting decision-making from a human being to an artificial agent such as a self-driving car raises several legal questions. Does the law permit artificial decision-making—or does it require human operators, at least in certain areas of law? How can artificial agents comply with legal norms such as road traffic regulations? And finally, what should self-driving cars do if they cannot avoid an accident and face tragic choices?

This chapter addresses these legal challenges posed by artificial decision-making. The legal questions are considered in an abstract manner, but with a view to German, U.S. and public international law, particularly human rights law. Other legal issues raised by self-driving cars (adaption

[5] This is at least a particular German concern. The Ethics Commission set up by the German Federal Ministry of Traffic in order to assess Autonomous Driving, for example, stated that outlawing human driving would interfere with the right of individual liberty (which also entailed the "pleasure of driving") and could not be justified by enhancement of safety alone (!), *Ethikkommission Automatisches und Vernetzes Fahren*, Bericht, Juni 2017, para. 5, p. 21 https://www.bmvi.de/SharedDocs/DE/Pressemitteilungen/2017/084-dobrindt-bericht-der-ethik-kommission.html.

[6] Not to mention the unpredictable consequences on the value of urban and rural land due to the cheaper costs of mobility, see Dirk Heinrichs "Autonomous Driving and Urban Land Use" in *Autonomous Driving, supra* note 3, 213.

[7] The terms algorithm and computer code or computer program will be used interchangeably throughout this chapter.

of international standards of road traffic law and product standards,[8] liability,[9] treatment of data used and generated by autonomous cars[10]) have to be analyzed elsewhere. After clarifying the relevant terms (section II), the chapter looks at the legal framework of artificial decision-making, in general (section III) and the legal problem of tragic choices, in particular (section IV).

II. TERMINOLOGY

Before delving into the legal issues, two terms merit clarification: autonomous driving and artificial intelligence.

A. Autonomous Driving

Autonomy as a legal or philosophical term is a very complex concept. In the context of self-driving cars, however, "autonomous" has a technical meaning which can be clearly defined. Since technological progress and commercial availability increase gradually, a widely-used terminology distinguishes different degrees of autonomy. Most writers refer either to a classification by the US-American National Highway Traffic Safety Administration (NHTSA) established in 2013 (ranging from level 0 to level 4) or to a classification by SAE International, a private association of engineers and related technical experts in the aerospace, automotive and commercial-vehicle industries, the former Society of Automotive Engineers, proposed in 2016 (Standard J3016, ranging from level 0 to 5)[11] (Table 9.1). The latter is more differentiated than the former and has, by now, also been adopted by the NHTSA.[12] As a consequence, it will be taken as a basis for this chapter as well.

[8] See, for example, Bryant Walker Smith, *Automated Vehicles are Probably Legal in the United States*.
1 Texas A&M University School of Law 411 (2014); see also below section III.B.2.

[9] See, for example, Melinda Florina Lohmann, *Liability Issues Concerning Self-Driving Vehicles* 7 European Journal of Risk Regulation 335 (2016); Daniel A. Crane, Kyle D. Logue and Bryce C. Pilz, *A Survey of the Legal Issues Arising From the Deployment of Autonomous and Connected Vehicles* 23 Michigan Telecommunications and Technology Law Review 191 (2017).

[10] *Cf.* Kai Rannenberg "Opportunities and Risks Associated with Collecting and Making Usable Additional Data" in *Autonomous Driving*, *supra* note 3, 497.

[11] Both classifications are reproduced, for example, in Dorothy J. Glancy, *Autonomous and Automated and Connected Cars—Oh My! First Generation Autonomous Cars in the Legal Ecosystem* 16 Minn. J.L. Sci. & Tech. 619, 630 (2015); a short description of the SAE classification can be found on the SAE's homepage www.sae.org/misc/pdfs/automated_driving.pdf.

[12] NHTSA *Federal Automated Vehicles Policy – Accelerating the Next Revolution in Roadway Safety September* 2016, p. 9, www.transportation.gov/AV.

Table 9.1. Levels Of Driving Automation As Defined By Sae International Standard J3016[13]

Human driver monitors the driving environment	
Level 0 No Automation	the full-time performance by the human driver of all aspects of the dynamic driving task,[14] even when enhanced by warning or intervention systems
Level 1 Driver Assistance	the driving mode[15]-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task
Level 2 Partial Automation	the driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task
Automated driving system (“system”) monitors the driving environment	
Level 3 Conditional Automation	the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene
Level 4 High Automation	the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene
Level 5 Full Automation	the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver

At level 0, solely the human driver is in charge of steering and acceleration/deceleration. At level 1, the car takes over a specific driving task, for

[13] Description taken from the SAE classification, supra note 11.

[14] Dynamic driving task includes the operational (steering, braking, accelerating, monitoring the vehicle and roadway) and tactical (responding to events, determining when to change lanes, turn, use signals, etc.) aspects of the driving task, but not the strategic (determining destinations and waypoints) aspect of the driving task.

[15] Driving mode is a type of driving scenario with characteristic dynamic driving task requirements (e.g., expressway merging, high speed cruising, low speed traffic jam, closed-campus operations, etc.).

example cruise control which keeps the car at a defined speed.[16] At level 2, the car can maintain two or more driving tasks—for example cruise control, automatic distance control and automated lane keeping—while the driver constantly monitors and controls the car. The most advanced automated cars which have been put up for sale until mid-2017 are level 2 cars, for example Tesla's Autopilot.[17] There is a decisive divide between level 2 and level 3: Cars from level 3 onwards drive by themselves. Level 3 cars, however, must still be constantly supervised by the human driver who has to be able to intervene promptly if a problem occurs. Level 4 cars do no longer need constant human supervision, but they are only capable to drive without human interference in common driving scenarios. Level 5 cars, finally, are able to drive without any human supervision or interference in all traffic or weather conditions in which a human driver could drive.

In this chapter, the term "autonomous" or "self-driving" car (used interchangeably) refers to level 3 to 5 cars, i.e. cars which are no longer driven by a human driver, but by a computer system. Thus, autonomous driving includes situations characterized by human supervision ("human *on* the loop" as opposed to "human *in* the loop") but also situations without any human role ("human *out of* the loop"), whereas the lower levels 1 and 2—characterized by a human driver in the loop—could be labelled "assisted" driving. Sometimes, "autonomy" is distinguished from "automation", emphasizing that autonomous systems are "intelligent" and capable to operate in an open surrounding whereas automated systems only fulfill simple, clearly defined tasks. In the literature on autonomous cars, both terms—automation and autonomy—can be found. Given the complicated tasks mastered by self-driving cars, it is, by all means, appropriate to speak of (technical) autonomy. Thus, driving decisions in autonomous cars are taken by the car (or rather the corresponding algorithm), not by a human driver.

B. Artificial Intelligence

There is less clarity in legal literature about the term "artificial intelligence" although self-driving cars and other artificial agents are commonly described as "intelligent" or "smart". Computer scientists distinguish four meanings of artificial intelligence. The Turing test approach, *first*, implies that a computer passes the intelligence test if a human interrogator, after posing questions and receiving answers from the computer, cannot tell whether the responses come from a human being or from a computer.[18] This test may be useful in other contexts, for example if an artificial personality is created to replace a human companion, but it does not make sense in the context of self-driving cars whose predominant task is to function

[16] *Cf.* Axel Davies "Everyone Wants a Level 5 Self-Driving Car—Here's What That Means" *Wired* August, 26 2016, www.wired.com/2016/08/self-driving-car-levels-sae-nhtsa/.

[17] *Id.*

[18] Stuart J. Russell and Peter Norvig *Artificial Intelligence. A Modern Approach* 3rd ed. (New Pearson 2010) p. 2.

reliably and safely. According to the *second* understanding artificial intelligence means that computers replicate the human mode of thinking.[19] This cognitive modelling approach (sometimes coined "strong artificial intelligence") is very ambitious as it requires deep insights into the working of the human mind and yet to be developed abilities of computer systems.[20] In our context, the second understanding can be set aside since self-driving cars work without cognitive modelling. The *third* approach equates intelligence with logical reasoning.[21] This approach, however, only works in formal settings where knowledge can be organized in logical notations, but not in real-life situations such as car driving which imply uncertainty, for example.[22]

According to the *fourth* approach, artificial intelligence describes rational behavior of artificial agents[23] or (to put it bluntly) intelligent outcomes.[24] It implies that these computer agents "operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue goals. A rational agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome."[25] This understanding is well-suited to capture the properties of autonomous agents ranging from mere computer programs (e.g. internet search engines, programs assessing the risk of recidivism in criminals) to physical machines like care robots, police robots, autonomous weapon systems or self-driving cars. Compared to the aforementioned approaches, it is more comprehensive than relying on logical reasoning alone (approach 3), less demanding than a replication of human cognition (approach 2) and it relies on similar computer abilities which would enable it to pass the Turing test (approach 1). In computer science, these abilities can be described namely as problem solving, knowledge representation, coping with uncertainty, learning, communication and robotics.[26] As a result, all autonomous cars are intelligent as they autonomously pursue goals (e.g. drive from A to B or find a free parking spot nearby) and thus autonomously reach intelligent outcomes.

Traditional legal concepts are not only challenged by artificial agents' autonomy, but more specifically by their learning abilities. Thus, a narrower understanding of artificial intelligence focusses on machine learning.[27] Machine learning implies that artificial agents improve their

[19] *Id.*, p. 3.

[20] *Id.*, p. 3. *Cf.* Harry Surden, "Autonomous Agents and Extension of Law" *Concurring Opinions*, February 16, 2012, https://concurringopinions.com/archives/2012/02/autonomous-agents-and-extension-of-law-policymakers-should-be-aware-of-technical-nuances.html.

[21] Russell and Norvig, *supra* note 18, p. 4.

[22] *Id.*

[23] *Id.*

[24] Surden, *supra* note 20.

[25] Russell and Norvig *supra* note 18, p. 4.

[26] *Id.*, chapters 3–25.

[27] *Cf.* Harry Surden, *Machine Learning and Law* 89 Washington Law Review 89 (2014); Amitai Etzioni and Oren Etzioni, *Keeping AI Legal* 19 Vanderbilt Journal of Entertainment and Technology Law 133 (2016).

behavior through experience, i.e. by training, not by following a fixed program.[28] Learning algorithms may, for example, recognize images of traffic signs after being fed with images depicting those signs.[29] The algorithms analyze data, detect patterns and build or refine models, mostly based on statistical calculations, in order to fulfill a given task. Well known applications of machine learning range from image and language recognition to spam filtering, language translation and the diagnosing of diseases or health risks.[30] Machine learning is particularly useful if it is too complicated to define all the steps necessary for fulfilling a given task (as in language translation, for example) or if these steps are unknown (as in characterizing the patterns of the risk of a certain disease, for example[31])—and if datasets exist or may be created. Learning algorithms might be also be used in self-driving cars, not only for image recognition, but also for airbag deployment or for finding the optimal path within a vehicle lane.[32] They could even enable a car to learn driving all by itself by observing human drivers.[33]

As a consequence, decision-making by artificial agents does not necessarily imply the use of machine learning. But, from a legal point of view, it will be relevant whether a certain feature of a self-driving car runs on a learning algorithm.

III. THE LEGAL FRAMEWORK FOR ARTIFICIAL DECISION-MAKING

How does artificial decision-making fit into the existing legal framework? Before characterizing artificial decision-making (subsection A), examining human operator requirements (subsection B) and addressing law compliance by artificial agents (subsection C), two aspects should be emphasized at the outset. First, the term "decision-making" by intelligent agents is not meant to carry the notion of human free will. Instead, it rests on the fact that these agents operate in an increasingly open surrounding. A driverless underground railway, for example, can operate at different speeds, but still has to stick to the railway tracks and normally follows a fixed schedule. Driving a car, however, implies choices regarding the speed and the location of the

[28] Russell and Norvig, *supra*note 18, chapters 18–21; Walther Wachenfeld and Hermann Winner, "Do Autonomous Vehicles Learn?" in *Autonomous Driving*, *supra* note 3, 451; Jason Tanz, "Soon We Won't Program Computers. We'll Train Them Like Dogs" *Wired*, June 2016, www.wired.com/2016/05/the-end-of-code/; Will Knight, "The Dark Secret at the Heart of AI" *MIT Technology Review* April 11, 2017 www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

[29] On this difficult task see Evan Ackerman, "Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms" IEEE Spectrum August 4, 2017, http://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms.

[30] Surden, *supra* note 27.

[31] Matthew Hutson, "Self-taught artificial intelligence beats doctors at predicting heart attacks" April 14, 2017, http://www.sciencemag.org/news/2017/04/self-taught-artificial-intelligence-beats-doctors-predicting-heart-attacks.

[32] Wachenfeld and Winner, *supra* note 28, p. 456.

[33] Knight, *supra* note 28 on the car by (chip maker) Nvidia.

car on the vast road network, including reacting to countless different traffic situations. Decision-making describes the task of picking one of the many options to act, depending on the circumstances of any given situation—regardless of whether the choice is taken by a human operator or a machine.

Second, law enjoys primacy over technology. Law determines the permissible operations of artificial agents—and not the other way round. This observation, trivial from a lawyer's perspective, is worth being recalled in the face of technological companies that claim leadership and preach technological solutions for the world's problems.[34] As a consequence, every legal norm affecting a self-driving car must be respected, technological capabilities or constraints notwithstanding.

A. Characteristics of Artificial Decision-Making

Human and artificial decision-making differ significantly: when facing the choice between stopping at a yellow traffic light or crossing the intersection, a human driver can, in principle, decide freely how to act even if the outcome will effectively be influenced by individual factors such as the driver's habits, her ability to assess the situation, her respect for traffic rules, or her emotional state. A self-driving car, on the other hand, is governed by algorithms which produce a definitive result for every situation depending on relevant factors such as the distance between the car and the intersection and the calculated breaking distance.

Thus, one of the most important advantages of autonomous cars is that they will unconditionally obey all legal norms duly reflected in the driving algorithms. Unlike human drivers who might speed, tailgate, take someone's right of way or jump a red light—due to emotions, fatigue, recklessness or outright egoism—autonomous cars can be programmed not violate traffic law. Proponents of other uses of artificial agents are also driven by the hope of fully law-abiding algorithms, for example of autonomous weapon systems which would never contravene the laws of armed conflict[35] or of automated law finding and law enforcement which would preclude the bias or the inaccuracy of a human judge or policeman.[36] In addition to the advantage of law-compliance unaffected by hu-

[34] For a critical account of this "solutionist" approach, see Evgeny Morosow, *To Save Everything Click Here* (Public Affairs 2013).

[35] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Ration 2009); Michael N. Schmitt and Jeffrey S. Thurnher, *"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict* 4 Harvard National Security Journal 2013, 231; Kenneth Anderson, Daniel Reisner and Matthew Waxman, *Adapting the Law of Armed Conflict to Autonomous Weapon Systems* 90 International Law Studies 386 (2014); Marco Sassóli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified* 90 International Law Studies 308 (2014).

[36] The use of a recidivism risk assessment algorithm in criminal sentencing has been affirmed by the Wisconsin Supreme Court *State v. Loomis* 881 N.W.2d 749 (Wis. 2016); some scholars endorse the idea that the process of law finding by the judiciary is supported or even replaced by algorithms, cf. the critical remarks by Kyriakos N. Kotsoglou, *Subsumtionsautomat 2.0. Über die (Un-)Möglichkeit ei-*

man traits and biases, artificial agents are also hoped to outperform humans in the knowledge of relevant facts and laws. Autonomous cars will recognize dangerous situations hidden to the human eye (for example a deer crossing a street at night) or analyze several driving options when faced with an unavoidable accident (choosing the least damaging outcome) and they will have access to traffic law in detail (including case law or laws of foreign countries on trips abroad). As a consequence, self-driving cars will be even better than humans at fulfilling general duties such as avoiding accidents or mitigating damages.

A closer look reveals, however, that law compliance is far more complex than the above discussion. If artificial agents operate in a surrounding defined by legal norms, several challenges arise.[37] *First*, artificial agents are incapable of law-finding. Law is made by humans and expressed in human language, it embodies human values and governs the life of human communities, it addresses human behavior and establishes rights and duties of human beings. As a consequence, law has traditionally been construed and applied exclusively by humans. For reasons of clarity and normativity, the terms "law" or "legal" norms, rules and principles should therefore remain limited to law in the traditional sense and should be distinguished from technical instructions (algorithms, computer code) governing the outputs of a computer system. Law's human essence does not preclude that it is translated into algorithms. But this is a challenging task (see below III.C.1.).

Second, artificial agents have difficulties in establishing specific facts, for example in assessing human behavior. Even the most advanced

ner Algorithmisierung der Rechtserzeugung Juristenzeitung 69 (2014), 451; in favor of these approaches Martin Fries, *Man versus Machine: Using Legal Tech to Optimize the Rule of Law*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2842726, p. 9; Anjanette H. Raymond and Scott J. Shackelford, *Technology, Ethics, and the Access to Justice: Should an Algorithm be Deciding Your Case?* 35 Michigan Journal of International Law 485 (2014); on possible techniques see Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro and Vasileios Lampos, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective* PeerJ Computer Science October 24, 2016, https://peerj.com/articles/cs-93/; on automated law enforcement (surveillance, analysis, action) see Woodrow Hartzog, Gregory Conti, John Nelson and Lisa A. Shay, *Inefficiently Automated Law Enforcement Michigan State Law Review* 1763 (2015).

[37] On the challenges of law-compliance by artificial agents see, for example, Ronald Leenes and Federica Lucivero, *Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design* 6 Law, Innovation and Technology 193 (2014); Henry Prakken and Giovanni Sartor, *Law and Logic: A Review from an Argumentation Perspective* 227 Artificial Intelligence 214 (2015); Amitai Etzioni and Oren Etzioni, *Designing AI Systems that Obey Our Laws and Values* 9(9) Communications of the ACM 29 (2016); Trevor Bench-Capon and Sanjay Modgil, *Norms and Value Based Reasoning: Justifying Compliance and Violation* 25 Artificial Intelligence and Law 29 (2017); on autonomous cars, in particular, see Henry Prakken, "On the Problem of Making Autonomous Vehicles Conform to Traffic Law" 25 *Artificial Intelligence and Law* 341 (2017), available at https://link.springer.com/journal/10506/onlineFirst/page/1; on legal automation, more generally, see Ugo Pagallo and Massimo Durante, *The Pros and Cons of Legal Automation and its Governance* 7 European Journal of Risk Regulation 323 (2016).

algorithms are unable to reliably predict human behavior in road traffic, e.g. the behavior of pedestrians[38] (and may still be fooled when reading traffic signs[39]), which prevents that traffic rules can be followed accordingly. Thus, the reliability of fact-finding is also a matter of legal concern (see below III.C.2.)

And *third*, algorithms—even though they might help to fight human biases—also embody biases. At first sight, self-driving cars seem less problematic than other uses of artificial intelligence, for example image recognition, online-advertising or criminal sanctioning which were found to produce racist outcomes.[40] But imagine, for example, that a self-driving car cannot avoid an accident and will either hit a bicyclist wearing a helmet or a bicyclist without a helmet.[41] Which collision should it choose? If the car is programmed to avoid the least vulnerable person, i.e. the bicyclist without a helmet, this will effectively punish those road users who protect themselves with a helmet. If the car is programmed, however, not to take headgear into account or to decide against road users not protecting themselves with a helmet, this might cause bigger damage and amounts to a bias against the unprotected road users. This shows that algorithms are never value free—they create biases of their own. As a consequence, a legal order must outlaw unacceptable tendencies and may regulate others. This will be illustrated by the problem of tragic choices, i.e. life-and-death decisions before an accident (see below section IV).

B. Human Operator Requirements

In some areas of law, artificial decision-making is implicitly or explicitly prohibited. There might be different reasons for such a human operator requirement—but the problem of artificial law compliance will often be one of them.

1. Human operator requirements in other areas of law

Human decision-making will regularly be required for the task of law-finding or exercising governmental authority. German civil and criminal procedural law explicitly states, for example, that the “court” (composed of one or several human judges) hears evidence and renders the decision, which

[38] Rodney Brooks, “The Big Problem With Self-Driving Cars is People” July 27, 2017 http://spectrum.ieee.org/transportation/self-driving/the-big-problem-with-selfdriving-cars-is-people.

[39] Note 29.

[40] On image recognition software tagging several African-Americans as gorillas *cf.* Megan Garcia “How to Keep Your AI From Turning Into a Racist Monster” *Wired* February 13, 2017, www.wired.com/2017/02/keep-ai-turning-racist-monster/; on online ads offering a person’s criminal record after an internet user has googled a black sounding name Latanya Sweeney, *Discrimination in Online Ad Delivery* 56 Communications of the ACM 44 (2013); on racial profiling Bernard E. Harcourt, *Against Prediction – Profiling, Policing and Punishing in an Actuarial Age* (University of Chicago Press 2006).

[41] On a similar example *cf.* Jeffrey K. Gurney *Crashing Into the Unknown: An Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law* 79 Albany Law Review 183, 197 (2015/16).

prevents outsourcing this task to algorithms.[42] Furthermore, legal norms may contain an implicit understanding that certain forms of decision-making are reserved to humans. German public law stipulates, for example, that public agencies exercise discretion when imposing a speeding fine. As a consequence automation is precluded and an individual decision by a (human) public servant is necessary.[43] Other legal requirements of individual decision-making also imply human assessment, e.g. an individual assessment of the defendant's guilt in criminal law,[44] an individual's right to be heard in administrative or judicial proceedings,[45] or the general prohibition of automated individual decision-making (including profiling) using personal data under EU data protection law.[46]

But human operator requirements can also be inferred from open-textured norms and the difficulty of artificial law compliance. Under the law of armed conflict, for example, a military attack is prohibited if it "may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated" (Article 51(5)(b) Additional Protocol I to the Geneva Conventions). It is not feasible to translate this assessment of proportionality into abstract rules of computer code, because it depends on multiple factors which are impossible to envisage and to evaluate in advance.[47] Thus, this provision of the law of armed conflict embodies a human operator requirement for weapon systems, at least if civilian losses have to be expected.

[42] *Cf.* § 286 German Code of Civil Procedure, § 261 German Code of Criminal Procedure.

[43] *Cf.* § 47(1)(1) *German Administrative Offences Act*; Higher Regional Court (OLG) Hamm, *Neue Juristische Wochenschrift* 1995, 2937; Higher Regional Court (OLG) Brandenburg, *Neue Zeitschrift für Strafrecht* 1996, 393; generally on the opposition of automation and discretion Danielle Keats Citron, *Technological Due Process* 85 Washington University Law Review 1249 (1303).

[44] According to § 46(1)(1) German Criminal Code, for example, the criminal sentence rests on the perpetrator's guilt. Furthermore, the principle of "nulla poena sine culpa" is also guaranteed by the constitution, see German Federal Constitutional Court, BVerfGE 20, 323, 331 (1966); on individualized considerations in sentencing under US law see Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination* 66 Stan. L. Rev. 803 (2014)—framing the problem mainly in terms of non-discrimination; Dawinder S. Sidhu. *Moneyball Sentencing* 56 Boston College L. Rev. 671 (2015); Andrea Roth, *Trial by Machine* 104 Georgetown Law Journal 1245, 1285 (2016).

[45] In German constitutional law, for example, the right to be heard is guaranteed in Art. 103(1) German Basic Law for judicial proceedings and flows from the rule of law (specified in § 28 German Administrative Procedure Act) for administrative proceedings; on the right to be heard under Art. 6 European Convention on Human Rights see European Court of Human Rights (ECHR) *Montovanelli/France* No. 21497/93 (1997), para. 33; *Goktepe/Belgium* No. 50372/99 (2005), para. 25; on the US perspective see Citron, *supra* note 43, p. 1305 (meaningful notice and opportunity to be heard), and the petition in *Loomis v. State of Wisconsin* Oct 5, 2016 (www.scotusblog.com/case-files/cases/loomis-v-wisconsin/) (due process rights in actuarial recidivism risk assessment when the algorithm is undisclosed and discriminatory).

[46] Art. 22 EU General Data Protection Regulation 2016/679.

[47] Noel. E. Sharkey, *The evitability of autonomous robot warfare* 94 International Review of the Red Cross 787, 789 (2012); Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal. Ethical, and Political Implications of*

2. *Human driver requirement*

Are human operators required under the law for self-driving cars, i.e. does the law demand a human driver? This issue is not only dealt with in national road traffic law, but also prescribed in two international conventions concluded under the auspices of the Economic and Social Council of the United Nations in order to facilitate international road traffic and to increase road safety through the adoption of uniform traffic rules. The Geneva Convention on Road Traffic (1949), ratified by nearly 100 state parties across the globe including the United States, Canada and other Commonwealth states,[48] stipulates in Article 8 (1): "Every vehicle or combination of vehicles proceeding as a unit shall have a driver." Similarly, the Vienna Convention on Road Traffic (1968), which was ratified by 65 predominantly European states[49] and replaces the Geneva Convention for its members,[50] demands in Article 8 (1): "1. Every moving vehicle or combination of vehicles shall have a driver." This raises the question if the term driver is confined to a human or can be understood to include artificial agent also.

Interestingly, US scholars seem to be more open for a reading which includes artificial agents[51] than German scholars.[52] The requirement of a "driver",[53] defined by the Conventions to be "a person" who drives a vehicle,[54] must traditionally be understood to characterize a natural person, not an artificial agent. (This understanding might change, of course, if artificial agents are granted legal personality in the future.) Systematical consideration supports this reading as both Conventions—the modernized Vienna

Autonomous Weapon Systems 47 Vanderbilt Journal of Transnational Law 1371, 1388 (2014).

[48] 125 UN Treaty Series 3; https://treaties.un.org/Pages/ViewDetailsV.aspx?src=TREATY&mtdsg_no=XI-B-1&chapter=11&Temp=mtdsg5&clang=_en.

[49] 1042 UN Treaty Series 17; https://treaties.un.org/Pages/ViewDetailsIII.aspx?src=TREATY&mtdsg_no=XI-B-19&chapter=11&Temp=mtdsg3&clang=_en.

[50] Art. 48 Vienna Convention on Road Traffic.

[51] Influential Bryant Walker Smith, *Automated Vehicles are Probably Legal in the United States*
1 Texas A&M University School of Law 411, 424 seqq. (2014).

[52] Lennart S. Lutz, *Autonome Fahrzeuge als rechtliche Herausforderung* Neue Juristische Wochenschrift 119, 123 (2015); Benjamin von Bodungen and Martin Hoffmann, *Das Wiener Übereinkommen über den Straßenverkehr und die Fahrzeugautomatisierung (Teil 2)* Straßenverkehrsrecht 93, 95 (2016); Antje von Ungern-Sternberg, "Völker- und europarechtliche Implikationen autonomen Fahrens" in: Bernd H. Oppermann and Jutta Stender-Vorwachs, *Autonomes Fahren* (C.H.Beck 2017) p. 293, 310 seq.

[53] On the rules of interpretation cf. Art. 31 Vienna Convention on the Law of Treaties 1969, UN Treaty Series 1155, I-18232.

[54] Art. 4(1) Geneva Convention: "'Driver' means any person who drives a vehicle, including cycles, or guides draught, pack or saddle animals or herds or flocks on a road, or who is in actual physical control of the same"
Art. 1(v) Vienna Convention: "'Driver' means any person who drives a motor vehicle or other vehicle (including a cycle), or who guides cattle, singly or in herds, or flocks, or draught, pack or saddle animals on a road".

Convention[55] more than the Geneva Convention[56]—establish duties of the driver which reflect human characteristics. Finally, from a teleological point of view, both Conventions aim at guaranteeing road traffic safety by imposing duties on a driver who is in control of the car[57] whereas upholding safety of self-driving cars would focus on the standards of these cars and their artificial agents. Thus, the Geneva and the Vienna Conventions (still) presuppose that a car is driven by a human, not by an artificial driver.[58]

Both Conventions could be amended, however, to allow for intelligent agents as car drivers.[59] Thus, the question ensues whether traffic law—like the law of armed conflict—is generally too open-textured to be followed by artificial agents. However, most norms of traffic law determine in a comparatively precise and comprehensive manner how a car can move (i.e. prescribing direction, speed, distance, right of way etc.). Even more general duties such as adjusting to traffic and whether conditions or showing mutual respect to other road users[60] can be categorized for typical situations (congestions, car accidents, approaching emergency cars, snow, glaze, fog or storm). And finally, in untypical and unpredictable situations, the car could be programmed to drive defensively, come to a standstill or demand that a human driver or remote operator takes over. As a consequence, road traffic law does not contain an implicit overall human operator requirement.

C. Law-Compliance by Artificial Agents

If artificial decision-making is legal in general, how can it be assured that artificial agents obey the law? This question is particularly important if artificial agents and humans operate simultaneously[61] in a densely regulated area of law (as opposed to, say, vacuum cleaning robots or internet search engines).

1. *Translating law into algorithm*

[55] For example Art. 8(3)(4) and (6) Vienna Convention demanding that drivers "shall possess the necessary physical and mental ability and be in a fit physical and mental condition to drive", "shall possess the knowledge and skill necessary for driving the vehicle" and "shall at all times minimize any activity other than driving" and—by domestic legislation—shall be prohibited to use "a hand-held phone while the vehicle is in motion".

[56] Art. 10 Geneva Convention demands, for example, that drivers "drive in a reasonable and prudent manner" and slow down "when visibility is not good".

[57] Art. 8(5) Geneva Convention; Art. 8(5) and Art. 13 Vienna Convention.

[58] Note that similar human driver requirements may also exist under national road traffic law. But not explicit requirements of a driver are found in the German Road Traffic Regulations or the US state laws governing road traffic; on the latter see Smith, *supra* note 51, 463.

[59] Generally by approval of a two-thirds majority of the state parties according to Art. 31(3) Geneva Convention; or even by silence of a two-thirds majority in response to an amendment proposal according to Art. 49(2) Vienna Convention.

[60] *Cf.* "Use of the road requires constant care and mutual respect", § 1(1) German Road Traffic Regulations.

[61] *Cf.* Glancy, *supra* note 11, 648.

It is clear, first, that the legal norms of road traffic law have to be translated into algorithms.[62] For practical reasons, the technical details of this translation will have to be developed by computer engineers. But the legal framework for this task will have to be specified by international and national regulation.

The translation can either proceed *top-down*, i.e. by deducing precise rules from general duties, or *bottom-up*, i.e. by teaching self-driving cars to model themselves on human drivers and to induce traffic rules from their behavior.[63] The top-down approach corresponds to the traditional legal technique of rule-making by administrative agencies or standard-setting private bodies. Under the top-down approach, the general duty to drive "at a careful and prudent speed not greater than nor less than is reasonable and proper, having due regard to the traffic, surface, and width of the highway and of any other condition existing at the time",[64] for example, would be translated into a specific maximum or minimum speed for a specific situation defined by speed limits, traffic conditions, weather and the like. Legal supervision is necessary in order to ensure that road traffic laws are interpreted faithfully and uniformly. As any other form of rule-making, the top-down approach has the advantage of being clear and predictable. But trying to anticipate and to regulate every imaginable situation in a comprehensive manner is very cumbersome. More importantly, rules are inflexible in untypical, unpredictable situations.

Thus, it may also be useful to let a self-driving car learn from human car drivers how to behave in these situations. It could learn, for example, to cross a solid line (which is generally prohibited) in order to cautiously circumnavigate an obstacle like a piece of dropped cargo—instead of bringing traffic to a standstill. Similarly, it could learn to align itself with the surrounding cars in order to create an emergency corridor (even if the emergency lane is established at the wrong place)—instead of obstructing the wrongly placed emergency corridor by following the legal rule. Bottom-up approaches are not alien to law-making and law-finding. Customary international law evolves from state practice and its acceptance as law,[65] and case law is developed by judicial decisions.[66] Thus, machine-learning, i.e. recognizing and reproducing patterns of (legal) behavior by self-driving cars, resembles the task of inducing legal rules from state behavior or judicial decisions. But unlike states which are entitled to develop customary international law and unlike courts which are authorized to clarify the law, road users do not, per se, qualify as a reliable source for lawful behavior, neither are they authorized to change the law. The examples show, on the other hand, that a flexible reading of legal rules helps to promote more general and possibly more important aims of traffic law

[62] *Cf.* Leenes and Lucivero, *supra* note 37; Prakken, *supra* note 37.

[63] On top-down and bottom-up approaches in general Wendell Wallach and Colin Allen, *Moral Machines – Teaching Robots Right From Wrong* (OUP 2009) chapters 6 and 7; Leenes and Lucivero, *supra* note 37, 4.4.

[64] Section 627(1) Michigan Vehicle Code Act 300 of 1949, to pick one of the US state codes at random.

[65] *Cf.* Art. 38(1)(b) Statute of the International Court of Justice.

[66] *Cf.* D. Neil MacCormick and Robert S. Summers, *Interpreting Precedents: A Comparative Study* (Dartmouth Publishing 1997).

such as keeping up the traffic flow or facilitating emergency operations. This could even include speeding, a very common violation of traffic law which should not be promoted generally, if it avoids or minimizes the risks of accidents, for example when a car merges onto the highway from an entrance ramp.[67]

As a consequence, the conditions of bottom-up rule-making by learning algorithms must be specified by law.[68] There are different modalities of machine-learning, allowing for a different degree of human input and control.[69] Imagine, for example, that the self-driving car would have to get clearance by a remotely operating legal officer in every new and untypical traffic situation before it was entitled to disrespect a rule of traffic law (and to reproduce this behavior in similar situations in the future). Contrast this with a process in which the self-driving car watches and imitates the behavior of the other cars, including reckless speeding and other violations of traffic law which are not justified by the circumstances. Faced with these extremes, upholding the primacy of law presupposes, at least, that machine-learning is supervised by humans and that any disregard for traffic rules can be legally justified. This would imply that self-driving cars are trained on classified sets of data which flag acceptable and inacceptable forms of driving, or that a human provides feedback as to the legality of learning results in the course of the learning process. Furthermore, every rule resulting from such a learning process must be identifiable (e.g. "It is permitted to cross a solid line in order to circumnavigate an obstacle if this does not endanger anyone") and explainable (e.g. by the importance of traffic flow). Any form of machine-learning which cannot (yet) explain its results would be precluded.[70]

2. *Establishing facts*

Law compliance by autonomous agents does not only require a correct understanding of the law, it also presupposes knowledge of the relevant facts. However, should law-finding and fact-finding, which are clearly distinguished in legal methodology, still be treated separately when artificial agents decide? Algorithms, after all, only deal with inputs and outputs. Yet, from a legal point of view, the distinction is still useful. Law-finding (e.g. "What is the speed limit?") is the preserve of lawyers, it follows legal methodology and may be classified as convincing/unconvincing or binding/non-binding. Facts, however (e.g. "At what speed does a car drive?"), are established with the help of non-legal disciplines, for example physics, follow the respective non-legal methodologies and may be classified as true or false. In reality, the difference is not quite as categorical: facts and

[67] Apparently, the google car is programmed to speed by up to 16km/h if this minimizes the risk of accidents http://www.reuters.com/article/us-google-driverless-idUSKBN0GH02P20140817.

[68] *Cf.* Benjamin I. Schimelman *How to Train a Criminal: Making Fully Autonomous Vehicles Safe for Humans* 49 Connecticut Law Review 327, 348 seq. (2016), advocating such a bottom-up approach without specifying the legal limits.

[69] On different forms of learning see Russell and Norvig *supra* note 18, p. 693 seq.; Wachenfeld and Winner, *supra* note 28, pp. 454–6.

[70] For a critical stance on "largely opaque and inscrutable" learning algorithms, see Knight, *supra* note 33; Etzioni and Etzioni, *supra* note 27, pp. 137–8.

laws are social constructs. Very often, facts may only be established by a certain degree of probability. Furthermore, the scientific methods rest on theories and models which might be falsified. Above all, fact-finding in the context of law is also governed by law, for example by procedural law directing the fact-finding of courts and administrative agencies. Nevertheless, the distinction between law and facts roughly separates the domain of lawyers and of other disciplines.

Thus, it is a technological question how reliably self-driving cars recognize their environment—and it is a legal question how reliable they should be when deployed on public streets. Defining the degree of reliability is a regulatory choice. But self-driving cars will arguably have to surpass human abilities in order to uphold the safety of road traffic and to promote confidence in autonomous driving. This is easier for some tasks (establishing weather conditions, recognizing objects beyond a human's field of vision, for example) than for others (recognizing human behavior, for example).[71] Would it have to be next to 100 percent, however, given that human fact-finding is not flawless either? Furthermore, a legal system cannot blindly trust in the reliability of algorithms. Instead, it must be able to comprehend how they function to assess their reliability. This may be illustrated by machine-learning in image recognition. Some forms of machine-learning, particularly learning in neuronal networks or "deep learning", are difficult to reproduce. The algorithm will learn to recognize cars or a traffic sign after analyzing a large set of data, but it cannot explain how the conclusions come about. Recent research has shown that this unsupervised self-learning process may lead to flawed outcomes. When researchers reproduced the results of two learning methods used to recognize images of a horse, they found that one algorithm based its results (understandably) on the contours in the picture. The other algorithm, however, drew upon the (purely coincidental) fact that the images of horses in the training set also showed a very small copyright sign and based its classification on this correlation. Thus, both algorithms would perform very differently in practice.[72] What is more, image recognition can even be subject to "adversarial attack". It was shown that small alterations to images of traffic signs (invisible to humans) significantly corrupted image recognition. This indicates how image recognition could be sabotaged not only by electronic image manipulations, but also by physically putting small stickers or a little bit of spray on street signs.[73]

IV. CRASH ALGORITHM'S TRAGIC CHOICES

Even if self-driving cars fully comply with road traffic law, there are further regulatory challenges due to the fact that artificial decision-making is never bias-free and sometimes even a matter of life and death. Self-driving cars, in particular, have to be programmed how to react when an accident is unavoidable, in other words how to choose among possible victims. Thus, crash algorithms will typically favor one group of possible victims over another one. Such a bias may be less dramatic if it protects more

[71] *Cf.* the need to recognize "military columns and motorised funeral processions" under Dutch traffic law, Leenes and Lucivero, *supra* note 37, fn. 79.

[72] Sibylle Anderl, *Frankfurter Allgemeine Zeitung* August 23, 2017 on these findings by the team of Sebastian Lapuschkin.

[73] Ackerman, *supra* note 29.

vulnerable cyclists and pedestrians at the expense of an armored car, but it is tragic if it involves decisions of life and death. This section will focus on these deadly choices which hopefully will help make the problem clear. It argues the tragic choices taken in crash algorithms are not merely a question of morality, but of law and should be regulated by government.

A. The Dilemma and ist Legal Dimension

Life-and-death dilemmas, i.e. situations in which every possible decision results in a tragic outcome, have been discussed by philosophers and criminal lawyers for a long time. The thought experiment ascribed to Carneades of Cyrene (taken up by Cicero and Kant) asks, for example, whether a shipwrecked sailor may push a fellow shipwrecked sailor from a plank carrying merely one person if this is the only means to survive.[74] The situation which most resembles the situation of a car accident is discussed in the "trolley"[75] or "switchman case" (as it is known in Germany),[76] developed with a view to abortion, ethical questions of medical progress and involvement in Nazi crimes:[77] it depicts a train which is about to crash into five men working on the railway tracks. Alternatively, the train could also be directed onto a different track, where only one person is working. It is certain that either the five workers or the one worker on the track will be killed by the collision. May the driver or the switchman decide to direct the trolley onto the other track, killing one, but saving five lives? Similar situations may arise in the course of road traffic. If a self-driving car is about to crash into five people crossing the road, should it swerve and drive onto the sidewalk, crashing into one pedestrian? More questions arise if more factors are taken into account. Is it permissible to save a child at the expense of an elderly person? Should a self-driving car be allowed to put the interest of its passenger first, even at the expense of many other road users? Should questions of responsibility be taken into account?

The choice of possible victims in an unavoidable accident raises important practical questions, even if the "trolley" or "switchman case" appear to be rather academic thought experiments.[78] It is true that self-driving cars will significantly decrease traffic accidents and that computer engineers are primarily concerned with enhancing the cars' safety by working on less spectacular questions such as the optimal speed or lane positioning. But accidents will not disappear as long as human road users

[74] Ulf Neumann, "Necessity and Duress" in: Markus D. Dubber and Tatjana Hörnle, *Oxford Handbook on Criminal Law* (OUP 2014) p. 583, 585.

[75] Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect* 5 Oxford Review 1, 3 (1967); Judith Jarvis Thomson, *The Trolley Problem* 94 Yale Law Journal 1395 (1985).

[76] Hans Welzel, *Zum Notstandsproblem* Zeitschrift für die Gesamte Strafrechtswissenschaft 47, 51 (1951).

[77] In post-war Germany, physicians who had reluctantly participated in the Nazi "Euthanasia" program argued to have prevented worse, *cf.* Welzel, *supra* note 76).

[78] See also Noah J. Goodall, "Machine Ethics and Automated Vehicles" Pre-print, published in Gereon Meyer and Sven Beiker, *Road Vehicle Automation* (Springer 2014) p. 93, available at http://dx.doi.org/10.1007/978-3-319-05990-7_9.

make mistakes and as long as technical failure by cars or infrastructure occurs. Thus, computer engineers are also designing "crash optimizing algorithms"[79] which determine a car's optimal behavior during an accident. These algorithms are more refined in predicting the future than the dilemma hypotheticals as they generate probabilities ("a 75 percent probability that a pedestrian be injured and a 60 percent probability that he be killed") instead of clear, but unrealistic results ("will die"). And they may take into account all the relevant aspects, proceeding from more common choices (e.g. whether to better collide with a big or a small car) to rare choices (e.g. whether to sacrifice the car's passenger in favor of a pedestrian).

Life-and-death algorithms of self-driving cars are currently being discussed as a matter of morality and ethics.[80] It will be argued, however, that they raise important legal questions and should be regulated by law. The classic thought experiments are used to consider reasons like necessity or duress which exclude criminal liability for decisions made in a dilemma, e.g. directing the trolley onto a certain track or pushing a shipwrecked off the plank.[81] The modern scenarios shift the focus: the tragic decision is no longer taken by a single person in a particular situation in a fraction of seconds, but it is determined by a crash optimizing algorithm developed in advance which establishes general rules for these tragic choices. Thus, it is now possible to create a meaningful legal framework for such an algorithm in advance instead of judging in retrospective how a human behaved in an extreme situation.

Different societies might favor different solutions for crash algorithms, just as they have developed different legal systems of criminal or torts law. Yet, they are constrained and guided by national and international human rights law, notably the right to life,[82] equality and non-discrimination[83] or human dignity.[84] It is true that these fundamental rights are (primarily) binding upon the state,[85] and not upon private manufacturers,

[79] *Cf.* Gurney, *supra* note 41.

[80] See generally Wallach and Allen, *supra* note 63, more specifically Goodall, *supra* note 78), and the literature below.

[81] Neumann, *supra* note 74, p. 585.

[82] Art. 2(2) German Basic Law; Am. 14 § 1 U.S. Constitution; Art. 2 European Convention on Human Rights (ECHR); Art. 4 American Convention on Human Rights (ACHR); Art. 6 International Covenant on Civil and Political Rights (ICCPR).

[83] Art. 3(1), 3 (3) German Basic Law (equality, non-discrimination); Am. 14 § 1 U.S. Constitution (equal protection of the law); Art. 14 ECHR (equal protection of convention rights); Art. 1, 24 ACHR (equal protection of convention rights, equal protection of the law); Art. 2(1), 26 ICCPR (equal protection of covenant rights, equality before the law, non-discrimination).

[84] Explicit guarantees are characteristic of younger human rights catalogues which reflect a history of inhuman and degrading treatment, *cf.* Art. 1(1) German Basic Law (human dignity); Art. 3 ECHR (prohibition of torture and inhuman or degrading treatment); Art. 5(2), 11 (1) ACHR (respect for the inherent dignity of the human person; honor and dignity); preamble ICCPR ("inherent dignity of the human person"); *cf.* chapt. 2, section 10 Constitution of the Republic of South Africa; *cf.* Niels Petersen, "Human Dignity, International Protection" *Max Planck Encyclopedia of Public International Law* 2012.

[85] Art. 1(3) German Basic Law; Art. 1 ECHR; Art. 1(1) ACHR; Art. 2(1) ICCPR; cf. *Jackson v. Metropolitan Edison Co.* 419 U.S. 345, 349 (1974).

computer engineers or users of self-driving cars. Nevertheless, many human rights regimes recognize that those particularly important rights also entail positive obligations,[86] i.e. duties to protect life,[87] equality,[88] and dignity[89] against violations by private parties. However vague these positive obligations are, they require at least some form of protection, particularly if grave violations are at stake which could easily be prevented by the state.[90] Thus, a regulation of crash algorithms can be understood to fulfill a positive human rights obligation (which would have to be established separately for every human rights regime). But even beyond the scope of such obligations, human rights considerations aptly characterize the specific legal interests at hand which speak in favor of regulation by the government.

It is useful to recall some further advantages of such a regulatory approach (the term "regulation" referring to governmental regulation only)[91]—as opposed to leaving the setting of crash algorithms up to the

[86] Walter Kälin and Jörg Künzli, *The Law of International Human Rights Protection* (OUP 2009) chapt. I.3.III.3; Olivier de Schutter, *International Human Rights Law* 2nd ed. (CUP, 2014) chapt. II.4.2.1.; see the obligation to "ensure" the convention/covenant rights (Art. 1 ECHR; Art. 1(1) ACHR; Art. 2 (1) ICCPR); this is understood to comprise all convention rights of the ACHR (Inter-American Court of Human Rights, judgment *Velásquez-Rodríguez v. Honduras (merits)* July 29, 1988, Series C No. 4, para. 166) and all convention rights of the ICCPR (Human Rights Committee *General Comment No. 31* May 26, 2004, CCPR/C/21/Rev.1/Add. 13, para. 8).

[87] For example German Federal Constitutional Court *Aviation Security Act* BVerfGE 115, 118, para. 120 (2006), available at http://www.bundesverfassungsgericht.de/entscheidungen/rs20060215_1bvr035705en.html (protection against terrorist acts); ECHR *Streletz/Germany* No. 34033/96 (2001), para. 86 (criminal protection); see also fn. 86.

[88] In Germany, this primarily results from an indirect horizontal effect of fundamental rights which implies that ordinary law is construed in accordance with specific antidiscrimination guarantees, see for example German Federal Constitutional Court, Neue Juristische Wochenschrift 2001, 2658 (needs of a disabled person have to be accommodated in a private tenancy); on positive obligations based on Art. 14 ECHR, see ECHR *Virabyan/Armenia* No. 40094/05 (2012), para. 218 (duty to investigate violent acts on political grounds); see also note 86.

[89] Based on Art. 1(1)(1) German Basic Law ("protect") see, for example German Federal Constitutional Court *Aviation Security Act* BVerfGE 115, 118, para. 121 (2006) (protection against "humiliation, branding, persecution, outlawing and similar actions by third parties"); ECHR *Pretty/United Kingdom* No. 2346/02 (2002), para. 51 ("ensure that individuals … are not subjected to torture or inhuman and degrading treatment or punishment … by private individuals"); Matthias Mahlmann, "Human Dignity and Autonomy in Modern Constitutional Orders" in: Michel Rosenfeld and András Sajó, *The Oxford Handbook of Comparative Constitutional Law* (OUP 2012) pp. 370, 384–5; see also note 86.

[90] Kälin and Künzli, *supra* note 86, chapt. I.3.III.3(c); de Schutter, *supra* note 86, chapt. II.4.2.1.

[91] On different concepts of "regulation" see Cary Coglianese and Evan Mendelson, "Meta-Regulation and Self-Regulation" in: Robert Baldwin, Martin Case and Martin Lodge, *The Oxford Handbook of Regulation* (OUP 2010) 146; Matthew T. Wansley, *Regulation of Emerging Risks* 69 Vanderbilt Law Review 401 (2016); Pagallo and Durante, *supra* note 37; Ronald Leenes, Erica Palmerini, Bert-Jaap Koops, Andrea Bertolini, Pericle Salvini and Federica Lucivero, R*egulatory challenges of robotics: some guidelines for addressing legal and ethical issues* 9 Law, Innovation and Technology 1 (2017).

car's manufacturer or the car's owner or driver ("self-regulation").[92] A regulation can counterbalance the conflicting interests of different road users (and of other relevant groups such as manufacturers and insurances) in a democratically accountable way. It creates clear and predictable rules and can therefore be taken into account by all road users. Finally, with a view to international traffic, those rules, or at least a certain set of rules, could be agreed upon at the international level by the parties to the Vienna or the Geneva Convention on Road Traffic (which would also be facilitated by drawing upon common human rights standards).

B. Specific Legal Questions

The specific legal questions open to such a regulation shall now be considered in turn.

1. *Death by algorithm and human dignity*

First of all, may life-and-death decisions be delegated from a human to an algorithm at all? Or does human dignity, one of the central tenets of the German or the South African constitution, for example,[93] forbid "death by algorithm" because it turns humans into mere items of a calculation (and would thus entail an obligation to outlaw those algorithms)? In the context of autonomous weapon systems, German and South African lawyers have objected to the use of lethal autonomous weapons on grounds of human dignity.[94] This reasoning could be extended to other life-and-death decisions by artificial agents. According to a common understanding, human dignity implies that all humans are of equal value and are treated as ends, not as means.[95] Algorithmic life-and-death decisions could be understood to legalize a situation which should never be considered legal, namely sacrificing some road users in order to save others. The legality of life-and-death algorithms could be seen as a legal acknowledgement that some lives are worthier than others.

However, systematic considerations show that human dignity does not prohibit algorithmic life-and-death decisions as such, as legal orders have already governed tragic choices without violating human dignity. Cri-

[92] *Cf.* Nick Belay, *Robot Ethics and Self-Driving Cars: How Ethical Determinations in Software Will Require a New Legal Framework* 40 The Journal of the Legal Profession 119, 122 seqq. (2015); Jan Gogoll and Julian F. Müller, *Autonomous Cars: In Favor of a Mandatory Ethics Setting* 23 Science and Engineering Ethics 681 (2017).

[93] *Cf.* note 84 and Mahlmann, *supra* note 89.

[94] Informal Meeting of Experts on Lethal Autonomous Weapons: Convention on Conventional Weapons Geneva: April 16, 2016, Panel on Human Rights and Lethal Autonomous Weapons Systems (LAWS), Comments by Christof Heyns, United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions, http://www.unog.ch/80256EDD006B8954/(httpAssets)/1869331AFF45728BC1257E2D0050EFE0/$file/2015_LAWS_MX_Heyns_Transcript.pdf, p. 5; Robin Geiss *Die völkerrechtliche Dimension autonomer Waffensysteme* (Friedrich-Ebert-Stiftung 2015) p. 8 seq.

[95] Mahlmann, *supra* note 89, p. 379; Petersen, *supra* note 84, para. 5.

minal law determines in a general and abstract manner whether dilemmatic choices (directing the trolley, pushing a shipwrecked off the plank, having an abortion) are punishable. Furthermore, life-and-death decisions are also regulated in other areas of law.[96] Vaccination is lawful and sometimes even mandatory despite the (extremely small) risk of causing death. Life-saving donor organs are distributed to the recipients according to certain criteria such as medical need, prospect of medical success or waiting lists spelled out in law or medical guidelines. Remember, finally, that the United States organized a lottery to draft soldiers in the Vietnam War. All of these regulations distribute risks of death in advance—the risk of being saved or killed by vaccination, by receiving a donor organ or waiting in vain, by evading conscription or by being drafted for a deadly war.

Neither of these legal arrangements is considered to violate human dignity. In the first situation of individual dilemmatic choices governed by criminal law, the law does not approve of the act of killing, for example, but shows understanding for a difficult personal decision. German criminal law, for example, attaches great importance to the distinction between the legality of an act and individual guilt. Thus, tragic choices might result in an illegal action (an act of killing), but might lack the element of individual guilt necessary for punishment.[97] In the second situation, law regulates tragic choices in advance—quite like crash optimizing algorithms do. The examples illustrate that this is necessary if a society wants to fight diseases by vaccination, to arrange for organ transplantations or to go to war, in other words if it advances important (life-saving) goals and accepts certain risks of death in exchange. As long as these risks are distributed fairly (e.g. among all by mandatory vaccination, by a waiting list for donor organs or by a randomized drafting procedure), this does not amount to a degrading treatment of those who will suffer from the distribution. This also applies for self-driving cars: if a society wishes to enjoy the benefits of self-driving cars—including a massive reduction of traffic fatalities—it will be necessary and legal to have algorithms envisaging critical life and death decisions. Human dignity is important in influencing these algorithms, but it does not prohibit them.

2. *Priorities: life, health, property*

After establishing that crash algorithms as such do not challenge human dignity, we can know address the necessary regulatory decisions. A regulation must, for a start, set a priority of legal values. It seems clear that a crash algorithm should prevent human fatalities first, human injuries second and damage to property third. This priority follows from the hierarchy of human rights (life, health, property) reflected in different standards for limiting these rights,[98] and from the gravity of criminal offenses

[96] From a theoretical point of view Guido Calabresi and Philip Bobbitt, *Tragic Choices* (Norton 1978).

[97] *Cf.* Thomas Rönnau *Grundwissen – Strafrecht: Übergesetzlicher entschuldigender Notstand* Juristische Schulung 113 (2017).

[98] Kälin and Künzli, *supra* note 86), chapt. I.3.III.2; de Schutter, *supra* note 86, chapt. II.3.2; II.3.3; Christian Tomuschat, *Human Rights* 3rd ed. (OUP 2014) chapt. 6 IV, V.

or torts (killing, bodily injury, damage to property) reflected in different criminal sanctions and damages.

This clear priority rule becomes less clear-cut if one takes a closer look. Could the prevention of a unique cultural site outweigh the risk of a minor bodily injury? If such considerations of proportionality are relevant in criminal law, torts law or human rights law,[99] they should also be legal in the context of autonomous cars. Which degree of probability is sufficient to establish danger to life? Is a 50 percent chance of a very serious bodily harm (e.g. brain damage) a worse scenario than a ten percent chance of killing somebody? These considerations are not as absurd as they might seem. The likelihood of an accident very often depends on the human behavior of other road users which cannot be predicted with accuracy, but only in terms of (rough) probabilities. Furthermore, the chances of being hurt or killed correspond to factors a self-driving car will soon be able to recognize, for example the gender and age of the victims and the mass of a car involved in an accident.[100] Taking account of probabilities is a common feature in law, for example medical law or police law. But defining precise thresholds and proportionalities is a delicate task. Finally: what role could other values play? Some societies, for example, attach particular importance to the protection of animals or of religious objects (or to both, imagine an algorithm designed to avoid holy cows strolling on Indian streets) based on constitutional principles or cultural and historical traditions. The examples show that even the seemingly simple task of defining the priority of legally accepted outcomes implies important choices. This, again, underlines the advantages of a democratic regulation.

3. *Personal charateristics and equal value of every life*

After emphasizing the value of human life, let's now assume that a deadly accident is unavoidable and that the car has two options, both of which will result in the death of a human. This life-versus-life scenario raises different questions which will now be addressed one by one. The tragic choice may, first of all, be guided by the personal characteristics of the possible victims. A crash algorithm could, for example, be designed to target an old person instead of a young child who has his whole life ahead of him. Other settings are easily imaginable: Feminists could favor women over men, racists whites over blacks, and utilitarians "useful" members of society over unfit people, homeless people, or criminals. Even if this result

[99] For example § 34 German Criminal Code; §§ 228, 904 German Civil Code; on proportionality in human rights law see note 98.

[100] Leonard Evans, *Death in Traffic: Why Are the Ethical Issues Ignored?* 2 Studies in Ethics, Law, and Technology 1, 8 (2000): "If one driver is a man, and the other a similar-age woman, the woman is 28% more likely to die. If one driver is age 20 and the other age 70, the older driver is three times as likely to die. If one driver is drunk and the other sober, the drunk is twice as likely to die (because alcohol affects many body organs, not just the brain). If one driver is traveling alone while the other has a passenger, the lone driver is 14% more likely to die than the accompanied driver, because the accompanied driver is in a vehicle heavier by the mass of its passenger."

may be supported by some people,[101] it is unacceptable from a human rights point of view. It is a central tenet of modern societies that every human life has equal value, regardless, in particular, of race, gender, age, or utility. This central tenet can be founded on human dignity which implies that every human being enjoys equal rights, equal respect and equal value.[102] Likewise, the principle can be directly based on equality and the guarantees of non-discrimination.[103]

What's more, this chapter claims that states even have a positive obligation to protect the equal value of every life against private forms of discrimination based on race, gender, age, or similar properties; in other words that states are obliged to outlaw corresponding settings that discriminate in crash algorithms. In some legal orders such as Germany's, this will, arguably, result from the prominent rank of human dignity.[104] But such a positive obligation can also be established under international human rights law given that the discriminatory crash algorithm would affect not only the right of human dignity or non-discrimination, but also the right to life. The paramount importance of these rights in international law can be illustrated by the fact that they are – at least partially – protected as *jus cogens*, i.e. as a peremptory norm of international law.[105] Furthermore, the principle of equal worth of every human being is reflected in many norms of national law or even private law, which illustrates that it is a well-established principle in many legal orders. In German criminal law, for example, killing an elderly person is considered no less heinous than killing a younger person who has his whole life ahead of him.[106] In the U.S., private discrimination is outlawed by specific anti-discrimination statutes, which compensates for a lack of horizontal effect or positive obligation entailed by fundamental rights.[107] At the level of professional standards, to provide a last example, the members of the Institute of Electrical and Electronics Engineers, in their Code of Ethics, have agreed, in any case, not to engage in "discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression".[108]

[101] A team of MIT presents a variety of dilemmatic decisions to internet users and asks them to choose the "lesser evil"; the suggested choices do, in fact, include targeting people according to their physical fitness, their profession and other characteristics (homelessness, criminality); http://moralmachine.mit.edu/.

[102] *Cf.* German Federal Constitutional Court *Aviation Security Act* BVerfGE 115, 118, para. 85, 121 (2006); Mahlmann, *supra* note 89, p. 380; Petersen, *supra* note 84, para. 29.

[103] Note 83; see also a combined reasoning based on both equality and dignity in Supreme Court of Canada *Law v. Canada* [1999] 1 S.C.R. 497.

[104] See again German Federal Constitutional Court *Aviation Security Act* BVerfGE 115, 118, para. 85 seq. (2006).

[105] It is generally accepted that the prohibition of slavery, genocide, arbitrary killing, racial discrimination, apartheid, and torture enjoy *jus cogens* status, Kälin and Künzli, *supra* note 86, chapt. I.2.III.2; de Schutter, *supra* note 86, chapt. I.3.4.2(b).

[106] German Federal Court of Justice *Decision*, August 11, 1995 – 2 StR 362/95 juris.

[107] For example the Civil Rights Act of 1964 prohibiting employee discrimination or harassment based on sex, race, color, religion, and national origin; the Age Discrimination in Employment Act 1967.

[108] http://www.ieee.org/about/corporate/governance/p7-8.html.

Crash algorithms which target victims according to their race, gender, age and other personal characteristics must therefore be outlawed. This means, in practice, that the tragic choice has to be decided by a random generator.[109] The procedure is fair by allocating the same risk of death to everyone. And randomization by computer algorithm is even more accurate than manual randomization. The Vietnam War Lottery, at least, allegedly did not produce truly random results—probably due to insufficient mixing of capsules.[110] "Random" does not mitigate the tragic situation, but it allows for a solution respecting the equality and dignity of the possible victims.

4. *Self-sacrifice and self-interest*

Having established the strict rule that every life is of equal value, we can now consider whether specific constellations might allow for specific solutions. Life-versus-life constellations will often affect the passenger of the self-driving car and another road user. Let's imagine, for example, that a car is on a narrow street at the cliffs or about to enter a tunnel, and that it can avoid a deadly collision with a pedestrian by driving off the cliff or into the wall of the tunnel which would result in the death of the car's passenger. May the crash algorithm choose to target another road user in order to avoid sacrificing its passenger?

Several aspects could be considered in support of egoistic settings. Online surveys show that people generally favor altruistic cars, but would prefer to buy a self-driving car with egoistic settings.[111] This (comprehensible) egoistic attitude will make acceptance of self-driving cars more difficult if they come with altruistic settings. From a legal point of view, it is argued, individuals must be allowed to opt for their own survival and must not be forced to ride in a self-sacrificing car.[112] And a utilitarian could ask whether the benefit of riding a self-driving car which significantly enhances overall road safety should not be promoted and rewarded at least by allowing for an egoist crash algorithm.

None of these arguments, however, justifies an exception from the principle of equal value of every life. It may be true that criminal law cannot demand self-sacrifice and must not sanction an egoistic individual operating as in the example of the plank of Carneades or before an unavoidable car accident.[113] However, once the state regulates this choice by mandatory randomization, as suggested, the passenger of a self-driving car is no longer in a position to choose between altruism or egoism. Instead, the autonomous driving mode relieves her not only from driving,

[109] *Cf*. Thomas Burri, Machine Learning and the Law: Five Theses, January 3, 2017 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2927625.

[110] Norton Starr, *Nonrandom Risk: The 1970 Draft Lottery* 5 Journal of Statistics Education (1997) https://ww2.amstat.org/publications/jse/v5n2/datasets.starr.html.

[111] Jean-Francois Bonnefon, Azim Shariff and Iyad Rahwan, *The Social Dilemma of Autonomous Vehicles* Science 352(6293) (2016).

[112] Philipp Weber, *Dilemmasituationen beim autonomen Fahren* Neue Zeitschrift für Verkehrsrecht 249, 253 (2016).

[113] *Cf.* Neumann, *supra* note 74; Rönnau, *supra* note 97; Weber, *supra* note 112).

but also from making dilemmatic crash decisions in advance, quite in the same way a passenger cannot control dilemmatic crash decisions on a train ride or a flight. A state is entitled and, based on positive human rights obligation, even bound to regulate accordingly. Neither can the other concerns—little acceptance or unfair burden of benefit—justify that some lives have a higher value than others and that passengers in a self-driving car are protected at the expense of all the other road users.

This example also illustrates the general need for a regulatory approach. Egoistic and altruistic decisions are not only a matter of life and death. From an insurer's point of view, for example, crash decisions will be assessed by the damage they cause. Thus, an egoistic crash setting chosen by a road user at the behest of her insurance company would minimize the damage to be covered by the company. This could produce undesirable results, for example, if it leads to a bigger overall damage or if it discriminates against certain road users, for example poorer people who drive cheaper cars. Autonomous cars will only be acceptable to society as a whole, if crashes are governed by fair, not by egoistic rules. Manufacturers, insurance, and road users, however, have no incentive to create and use altruistic settings if their competitors and fellow road users will not do the same—which requires mandatory settings.[114]

5. Numbers

A further problem is raised by quantitative considerations. Let's come back to the example of a car which will either crash into a pedestrian (killing *one*) or into another car with five passengers (killing *five*). Could or should the crash algorithm favor killing one over killing five? From an ethical point of view, the comparable decision in the trolley or switchman scenario is held to be morally permissible or even mandatory.[115] Some philosophers focus on distinguishing the trolley/switchman scenario from other—purposely absurd—cases where choosing to kill one person to save five is clearly unmoral: killing a person to enable organ transplants which would save five lives, pushing an overweight bystander onto the tracks to stop the trolley which is heading in the direction of the five workers[116] or blowing up an overweight man who is stuck in the mouth of a cave, trapping the exit of his fellow potholers who are facing death by rising flood waters in the cave.[117] It is explained that actively killing one is worse than letting five die,[118] or that directly violating the right of life is worse than doing something which is not in itself a violation of a right like turning the trolley.[119] To both aspects one could add a third one, namely that it is only morally acceptable to make such a choice if it affects people who are already in imminent danger of death—and not third parties. Once these groups who are threatened by death are established, i.e. two groups of workers on two different tracks or two groups of road users both facing

[114] Gogoll and Müller, *supra* note 92.

[115] Welzel, *supra* note 76, 51; Foot, *supra* note 75, 3; Thomson, *supra* note 75).

[116] Thomson, *supra* note 75, 1409.

[117] Foot, *supra* note 75, 2.

[118] *Id.*, 4.

[119] Thomson, *supra* note 75, 1403.

death by a possible collision, it is the simple idea of reducing the overall death toll which legitimizes killing one instead of five.[120]

Other philosophers emphasize that a crash algorithm decides tragic choices in advance and claim that a quantitative solution is therefore not only compatible with a consequentialist, but also with a deontological approach.[121] In contrast to consequentialism, deontology judges the morality of choices not by their effects, but by their conformity with a moral norm, for example the Kantian injunction against using others as mere means to one's end.[122] A quantitative approach can be easily defended on grounds of consequentialism as it saves more lives. From a deontological point of view, however, strict rules such as the prohibition of torture have to be obeyed regardless of the costs. The philosophers seem to proceed from the assumption that a moral decision has to duly respect the interests of the persons involved.[123] They then argue that programming an algorithm to minimize the death toll must be judged ex ante, i.e. without knowing the actual victims. Since the algorithm equally reduces everybody’s chance of becoming a victim it duly respects everybody’s interests.[124] It is certainly true that decisions taken in advance may be judged differently than decisions taken in an extreme situation. But this argument does not in itself help to overcome strict deontological rules.

From a legal point of view, i.e. by human rights standards, the state may (but need not) mandate quantitative decision-making by crash algorithms.[125] First, such a regulation does not violate the right to life. By regulating road traffic, the state does not actively interfere with the right to life.[126] Instead, the state fulfills its positive obligation to protect the life of road users by creating and enforcing traffic laws or car safety standards. Similarly, prescribing a death-toll-minimizing crash algorithm would protect the life of many potential death victims and thereby fulfill a state’s positive obligation towards life. Second, such a regulation does not violate human dignity or equality, i.e. the principle of the equal value of every human, even if it might seem to value the lives of the survivors more than the lives of the victims. But faced with the unavoidable death of many, it is fair rule to save as many lives as possible. It does not discriminate against

[120] Foot, *supra* note 75, 5.

[121] Alexander Hevelke and Julian Nida-Rümelin, *Selbstfahrende Autos und Trolley-Probleme: Zum Aufrechnen von Menschenleben im Falle unausweichlicher Unfälle* 19 Jahrbuch für Wissenschaft und Ethik 5 (2015).

[122] Larry Alexander and Michael Moore, “Deontological Ethics” in: Edward N. Zalta, *The Stanford Encyclopedia of Philosophy* Winter 2016 https://plato.stanford.edu/archives/win2016/entries/ethics-deontological/>, paras 1, 2, 2.4.

[123] Hevelke and Nida-Rümelin, *supra* note 121, 11.

[124] *Id*., 12.

[125] The German governmental “Ethics Commission on Automated and Connected Driving” was rather cryptic in its findings: It condemned sacrificing one person to save several others, but sanctioned minimizing the death toll if people were in imminent danger of death, *Ethikkommission Automatisches und Vernetzes Fahren*, *supra* note 5, para. 1.6, p. 18.

[126] In contrast to a law authorizing to shoot down an aircraft in case of a terrorist attack, which was held to violate human dignity, German Federal Constitutional Court *Aviation Security* Act BVerfGE 115, 118 paras. 84 seq. (2006).

certain groups of people defined by personal characteristics but relies purely on the (coincidental) fact whether a person belongs to a bigger or a smaller group of possible victims. Similar quantitative considerations are well established in other areas of the law. In criminal law, manslaughter of five is punished more severely than the killing of one. In situations of an emergency, public authorities and emergency forces have a discretion to allocate resources so as to save as many lives as possible. In the law of armed conflict, finally, legality of a military attack is determined, among other factors, by the number of expected civil casualties.[127] One should be much more cautious, however, to claim that a state is obliged to prescribe a quantitative approach. Positive human rights obligations such as the duty to protect life entail only a few very specific obligations, and leave the means of protection, generally, up to a state.[128] Furthermore, it is difficult to imagine that self-driving cars (in contrast to emergency forces, for example) face a decision involving huge differences in fatalities which would clearly speak in favor of a quantitative solution, for example saving 100 people instead of one person. Thus, the state may prescribe a death-toll-minimizing crash algorithm, but is not obliged to do so.

6. *Further regulatory choices: areas of risk and responsibility*

Finally, there are other possibilities which could be considered when regulating crash algorithms. This section will present two regulatory choices which would modify an approach based on randomization and death toll minimization.

The regulator could, *first*, specify the risk of being the victim of a car accident. As I have emphasized above, all the possible victims in the trolley scenario are in imminent danger of death—as opposed to third parties, for example a healthy person, who must not be killed even if his organs would save many lives. A closer look reveals, however, that being in imminent danger of death is a matter of legal assessment and thus open to regulatory definition. In the trolley case, for example, being killed by an uncontrollable train could be considered a) a general risk of life which affects anybody within reach of a train (including bystanders who would be killed by derailing), b) a risk affecting anybody who is working or otherwise present on railway tracks, or c) a risk affecting only those persons working or otherwise present on particular railway tracks, i.e. those tracks in whose direction the train is actually heading. A moral position which allows to divert or even derail the train to prevent worse, rests on the presumption that the risk of being killed by a train is a risk according to b) or even a). Being deliberately killed in order to donate organs, however, is not acknowledged as a risk of life in a society committed to the right to life. With respect to road traffic, one can equally distinguish different understandings of the risk to be killed by a car: this risk could be considered, for example, to be a) a general risk affecting anybody within reach of an unstoppable car (even on private premises, for example), b) a risk affecting

[127] See above III.B.1 on Art. 51(5)(b) Additional Protocol I to the Geneva Conventions.

[128] Kälin and Künzli, *supra* note 86, chapt. I.3.III.3(c); de Schutter, *supra* note 86, chapt. II.4.2.1.

all road users (including pedestrians on sidewalks or cyclists on cycle lanes), c) a risk affecting all persons who are present on the roads or d) a risk affecting only those persons who are present on the very lane the unstoppable car is driving on.

Even if these risks have never been defined so far, it is possible for the legislator to do so now that crash algorithms can be programmed accordingly. The legislator may stipulate, for example, that in a life-and-death scenario, a self-driving car may never choose to leave the public traffic space, or to leave the roads, or to leave a particular traffic lane—and the groups of possible victims would vary accordingly. The regulatory power flows from the legislator's power to distribute the risks associated with modern technologies in general and road traffic in particular. Such a regulation does not challenge the equal value of those persons in risk of being killed by a car: it does not weigh up life, but defines the rules of road traffic in the same way other traffic rules do. This also seems to be the position of those who claim that "third parties" should not bear the burden of autonomous driving.[129] It is a discretionary task however, to define who exactly counts as a "third party" and who does not. General principles of risk allocations could guide the legislator. One could argue, for example, that only motor vehicles contribute to the risks of car accidents and that only their drivers or passengers profit from a particular fast and convenient form of mobility (and thus exclude that pedestrians or cyclists become deliberate victims of a car accident). But one could also argue that society as a whole causes the risk of and profits from modern forms of mobility which are indispensable for the distribution of goods and services. A more convincing reason for limiting crash options is predictability. Imagine that a self-driving car cannot avoid an accident and has the option to collide either with a car on its traffic lane or on the opposite traffic lane. Having two options allows for flexibility (and for minimizing the overall death toll, for example), but reducing the options by outlawing that a car leaves its own lane (leaves the road, leaves the public traffic area) makes car accidents predictable and enables manufacturers and road users to act accordingly.[130]

A *second* regulatory choice concerns responsibility. Resorting to responsibility in the design of crash algorithms could mean, for example, that a self-driving car prefers a collision with a careless jaywalker over a collision with another pedestrian. Such a responsibility-based approach offers an incentive for responsible behavior. Responsibility is also a commonly accepted criterion for the allocation of risks from a legal point of view. In equality law, treating people differently according to their behavior is less suspicious and more easily justifiable than treating them differently based on personal characteristics. In police law, a hostage taker or a terrorist might even be actively killed as a matter of last resort in order

[129] *Ethikkommission Automatisches und Vernetzes Fahren*, *supra* note 5, para. 9, p. 11 "Those involved in creating the risk of mobility are not entitled to sacrifice third parties" (my translation); generally Alexander Hevelke and Julian Nida-Rümelin, *Ethische Fragen zum Verhalten selbstfahrender Autos bei unausweichlichen Unfällen: Der Schutz von Unbeteiligten* 69 Zeitschrift für philosophische Forschung 217, 217 (2015).

[130] *Cf.* Hevelke and Nida-Rümelin, *supra* note 129, 222.

to save the lives of the hostages or the possible victims of terrorism. Even the broad notion of human dignity developed by the German Federal Constitutional Court does not imply otherwise. The Court held, in the context of antiterrorism efforts, that shooting down an airplane hijacked by terrorists would violate the dignity of the hijacked air passengers (as they would be "used as means to save others"), but specified that shooting down the terrorists themselves would not qualify as a violation.[131] In this constellation, human dignity is apparently not at stake with respect to the perpetrators, which might be explained by the fact that they are considered self-governed agents who knowingly put their lives at risk, and not instruments of state action. Thus, if considerations of responsibility might even justify lethal forms of law enforcement, it will also be admissible to take responsibility into account when designing crash algorithms.

Yet, a closer look reveals that it is very difficult to allocate the burden of a car accident in accordance with responsibility. In many situations, responsibility for an imminent accident is unclear and will only be established afterward. Some accidents will be caused by a malfunction of vehicles or road infrastructure which is difficult to attribute to possible victims. Some of the road users who cause an accident such as children will not be considered responsible. Finally, even if a particular road user is definitively responsible for a dangerous situation, it will not always be possible to react in a way which sanctions him or him alone. Imagine, for example, that a bus driver has fallen asleep which provokes the bus to drive on the opposite lane. Crashing into the bus would not only risk his death, but also the death of many innocent passengers (let alone the death of the passengers of the self-driving car on the opposite lane). As a consequence, the regulator will have to examine if and to what extent a responsibility-based approach is feasible at all.

V. CONCLUSION

Artificial intelligence results in decision-making by artificial agents replacing human decision-making. This chapter has examined the challenges raised by this development with respect to the densely regulated area of traffic law in which self-driving cars and human drivers operate simultaneously. As opposed to other, more open-textured areas of law, road traffic law is generally suited for the operation of artificial agents (even if the requirement of a human "driver" in international conventions still needs to be amended). Thus, it is now up to the legislator and to other regulatory bodies to clarify the legal framework. In order to ensure law compliance

[131] German Federal Constitutional Court *Aviation Security* Act BVerfGE 115, 118 (2006), para. 124, on the hijacked passengers: "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 ...". The Court's solution may be understood to express caution when assessing and responding to terrorist threats given that it is not easy to establish, for example, whether the hijackers are effectively about to fly into a skyscraper, a nuclear plant or soccer stadium. It is, however, less convincing to prohibit any form of death toll minimization on grounds of human dignity if one group of victims cannot be saved, anyhow.

by self-driving cars, the regulator has to supervise how norms of traffic law are translated into computer code and has to set standards of reliability for artificial fact-finding. Machine learning, which will probably be an important method not only for fact-finding, but also for meaningful law compliance, is not incompatible with the primacy of law if the process is supervised and if its results are explainable and justifiable. Furthermore, a regulator has to deal with the biases created by artificial decision-making as this chapter has illustrated with respect to tragic life-and-death decisions of crash algorithms. It is argued that crash algorithms raise legal, not only moral, questions and should be regulated by law. In doing so, the regulator fulfills a positive obligation flowing from human rights, i.e. the right to life, equality, and human dignity, which should also generally guide the regulatory choices. More precisely, crash algorithms, which do not, as such, violate human dignity, will have to reflect the priorities of a legal order and must not use personal characteristics such as race, gender, or age, to choose between potential victims of an accident. The regulator may, however, prescribe death-toll minimization, specify areas of risks or resort to responsibility as a relevant criterion for those tragic decisions. On the whole, the regulation of artificial decision-making by self-driving cars will probably be only one example of how to organize the coexistence of humans and artificial agents. Many others will follow.

SOURCES

Cases

- European Court of Human Rights (ECHR) *Montovanelli/France* No. 21497/93 (1997)

- ECHR *Streletz/Germany* No. 34033/96 (2001)

- ECHR *Pretty/United Kingdom* No. 2346/02 (2002)

- ECHR *Goktepe/Belgium* No. 50372/99 (2005)

- ECHR *Virabyan/Armenia* No. 40094/05 (2012)

- Inter-American Court of Human Rights *Velásquez-Rodríguez v. Honduras* (merits) (1988) Series C No. 4

- Supreme Court of the United States *Jackson v. Metropolitan Edison Co.* 419 U.S. 345 (1974)

- Wisconsin Supreme Court *State v. Loomis* 881 N.W.2d 749 (2016)

- Supreme Court of Canada *Law v. Canada* 1 S.C.R. 497 (1999)

- German Federal Constitutional Court BVerfGE 20, 323 (1966)

- German Federal Constitutional Court *Aviation Security Act* BVerfGE 115, 118 (2006)

- German Federal Court of Justice, 2 StR 362/95 (1995) juris

- Higher Regional Court Hamm, Neue Juristische Wochenschrift 1995, 2937

- Higher Regional Court Brandenburg, Neue Zeitschrift für Strafrecht 1996, 393

International and European Law

- Statute of the International Court of Justice (1949)

- Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts 1125 UNTS 3

- Geneva Convention on Road Traffic (1949) 125 UNTS 3

- European Convention of Human Rights (1950) 213 UNTS 221

- International Covenant on Civil and Political Rights (1966) 999 UNTS 171

- Vienna Convention on Road Traffic (1968) 1042 UNTS 17

- American Convention on Human Rights (1969) 1144 UNTS 123

- Vienna Convention on the Law of Treaties (1969) 1155 UNTS 331

- European Union General Data Protection Regulation (EU) 2016/679

National Law

- Constitution of the United States

- Michigan Vehicle Code

- German Basic Law

- German Criminal Code

- German Code of Civil Procedure

- German Administrative Offences Act

- German Administrative Procedure Act