Digital challenges to personal autonomy for German criminal law
– Do we need a "new" criminal law for autonomous robots? –

Dr. Johanna Göhler, LL.M. (Yale)[1]

Abstract

Should we hold an autonomous robot that has caused harm to others criminally liable? The emerging technology of autonomous robots touches upon our perception of the foundations of criminal liability in an unprecedented way. Essentially, it raises the question whether we have to redefine personal autonomy as a foundational requirement of criminal liability in order to tackle digital developments. This paper analyses the application of conventional doctrines of criminal liability to the scenario of an autonomous robot harming others. Based on the analysis, the paper proposes a partial shift of risks associated with the use of robots from the individual user onto society by limiting individual criminal liability in order to reconcile the competing interests. The paper further shows why holding robots criminally liable is, at least as things stand today, not a feasible solution. Instead, the paper suggests alternative legal measures that might be better suited to address the challenges and close a potentially emerging liability gap than the (exclusive) reference to criminal law.

I. Introduction

Digitalization poses manifold challenges to German criminal law. Such challenges include the formulation of new offences targeting the destruction and manipulation of data,[2] the effects that a merger of human tissue and nano electronic devices can have on individual culpability, e.g. in the case of deep-brain-stimulators,[3] and the potential criminalization of a digital enhancement of human bodies[4].

The digital development calling most urgently for legal discourse, however, is the evolution of autonomous robots. This urgency reflects in numerous political initiatives on the European and national levels. In February of 2017, the European Parliament, for example, issued a resolution to the European Commission, pathetically requesting the latter to take action in respect of the legal and ethical issues raised by the developments in robotics and artificial intelligence: "whereas now that humankind stands on the threshold of an era when ever more sophisticated robots, bots, androids and other manifestations of artificial intelligence ("AI")

[1] The author thanks the participants of the German-Italian symposium „Personal autonomy in the digital world", conducted by the *Associazione italiana di diritto comparato* and the *Deutsche Gesellschaft für Rechtsvergleichung* at the University of Ferrara, Italy, in October of 2017, for their valuable comments. This paper is based on the talk given by the author during that symposium and dates from December 2017.

[2] Cf. §§ 303 a, b German Penal Code („Strafgesetzbuch").

[3] See in more depth *Beck*, Technisierung des Menschen – Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept „Verantwortung", in: Gruber/Bung/Ziemann, Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft (2014), p. 173, 175 seq.; *Beck*, Roboter und Cyborgs – erobern sie unsere Welt?, in: Beck, Jenseits von Mensch und Maschine (2012), p. 9, 10 seq.; *Beck*, JR 2009, 225, 228 seq.

[4] See, e.g., *Brunhöber*, Individuelle Autonomie und Technik im Körper, in: Beck, Jenseits von Mensch und Maschine (2012), p. 77 seq.

seem to be poised to unleash a new industrial revolution, which is likely to leave no stratum of society untouched, it is vitally important for the legislature to consider its legal and ethical implications and effects, without stifling innovation".[5] On a similar vein, the German Federal Minister of Transport and Digital Infrastructure set up an Ethics Commission to receive the supposedly first guidelines in the world for automated driving.[6] The Ethics Commission published its answers to new legal and ethical challenges raised by human-machine interaction in June 2017.[7] In its guidelines, the Commission established 20 principles aiming to reconcile the competing interests in safety, human dignity, personal freedom of choice, and data autonomy.[8]

The political initiatives, so far, primarily target the adoption of private law rules. However, the digital revolution of autonomous robots raises questions for criminal law, too. This becomes particularly obvious when a robot harms a third party. Traditionally, in German criminal law, a human being is held liable for harm if she has culpably caused the harm by personal misconduct. With robots becoming more and more autonomous, this traditional approach becomes increasingly more difficult to apply. Therefore, this piece examines whether personal autonomy as a foundational requirement of criminal liability under German criminal law has to be redefined in order to tackle digital developments.

To set the scene for the analysis, the piece starts by defining the terms *personal autonomy* and *autonomous robots*. Thereafter, a case study is used to illustrate the (legal) peculiarities of an autonomous robot harming a third party. Based on this case study, the paper analyses and challenges the application of conventional doctrines of criminal negligence liability to the scenario that an autonomous robot has caused harm. The analysis concludes with suggestions how to reconcile the competing interests involved. Essentially, it proposes a partial shift of risks associated with the use of autonomous robots from the individual user onto society at large by limiting individual criminal liability. Attending to the fact that this proposal as well as other factors idiosyncratic to the field of robotics may potentially result in a liability gap,

[5] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), P8_TA(2017)0051, para. B. The resolution is annexed by draft civil law rules on robotics. The Commission responded rather cautiously on May 16, 2017, see Follow up to the European Parliament resolution of 16 February 2017 on
civil law rules on robotics.

[6] See *Dobrinth*, quoted in press release 084/2017 of the Federal Ministry of Transport and Digital Infrastructure, *available at* https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/084-ethic-commission-report-automated-driving.html (last accessed Oct. 23, 2017).

[7] Federal Ministry of Transport and Digital Infrastructure, Ethics Commission. Automated and Connected Driving, Report, June 2017, *available at* https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile (last accessed Oct. 23, 2017) (in the following: Commission, Report).

[8] Criminal law scholars had primarily awaited the Commission's answer to the well-known trolley-problem in modern automated driving contexts; see on this issue *Weber*, NZV 2016, 249, 251 ff.; *Weigend*, ZIS 2017, 599 seq. However, the Commission disappointed those who had expected clear-cut standards that could be easily transposed into a digital code governing the car's "decision" in dilemmatic situations, i.e. situations „in which an automated vehicle has to "decide" which of two evils, between which there can be no trade-off, it necessarily has to perform" Commission, Report, principle 5. Instead, the Commission shifted the primary responsibility on computer scientists and engineers by requiring them to design a technology that avoids such dilemmatic situations in the first place, *id.* principle 5. For the occasion that a hazardous situation is technically unavoidable, the Commission confirmed the validity of some ethical principles known from analog or non-digital context, see *id.* principles 7, 9. Beyond that, the Commission stated that legal judgments governing the *ex post* assessment of the individual guilt, respectively excuse of an individual driver in a dilemmatic situation could not be readily transformed into abstract/general *ex ante* appraisals and thus also not into corresponding programming activities, *id.* principle 8.

the paper subsequently scrutinizes recent suggestions to hold robots themselves criminally liable. The careful assessment, however, reveals that the relevant proposals have not yet succeeded in vindicating that autonomous robots would possess personal autonomy as necessary prerequisite of culpability and hence criminal responsibility and punishment. These findings lead to the initially phrased question whether personal autonomy as a foundational requirement of criminal liability needs to be redefined in order to tackle digital challenges. In response, the last part of the paper develops critical arguments derived from constitutional law, penal theory, legal philosophy, and information technology which should be acknowledged when contemplating to adapt criminal law theory in order to hold autonomous robots liable. The paper concludes by highlighting some alternative measures that might be better suited to address developments raised by robotics than the (exclusive) reference to criminal liability.

II. Definitions and concepts

The terms *personal autonomy* and *autonomous robots* can mean different things in different contexts. To set the scene for the following analysis, it is therefore imperative to define the meaning of these terms for the given context.

1. Personal autonomy

Traditionally, German criminal law is founded on the concept that only human beings[9] are suitable subjects of criminal responsibility because only human beings posses personal autonomy. Personal autonomy encompasses that a person has (1) the moral ability to distinguish between the right and the wrong, (2) the ability and freedom to decide in favor of the right and against the wrong, and (3) the ability to adapt her behavior according to this decision. Hence, in the present context, personal autonomy is used in the sense of *free and responsible moral self-determination* as it has been coined by the German Federal Court of Justice in 1952.[10] As such, personal autonomy is a prerequisite for the ascription of personal culpability, i.e., guilt. Guilt, in turn, is a condition of criminal responsibility and punishment.

2. Autonomous robots

In this paper, the term *autonomous*[11] *robot* is used to describe a structure with the following characteristics: it has a physical embodiment that bestows on it existence in the real world as opposed to only in a virtual world,[12] it is not living in a biological sense, it is aware of its

[9] In earlier times, non-human entities, such as animals, have been held criminally liable, too, see e.g., *Gleß/ Weigend*, ZStW 126 (2014), 561, 566 seq. In the context of robotics, proponents and opponents of criminal liability of robots draw arguments from this animal rights and liability discourse; see *Hildebrandt*, Criminal Liability and "Smart" Environments, in: Duff/Green, Philosophical Foundations of Criminal Law (2011), p. 507, 515 seq.

[10] BGHSt 2, 194, 200 seq., calling it "freie, verantwortliche, sittliche Selbstbestimmung".

[11] This paper uses the term *autonomy* in a sense of technological autonomy, not moral consciousness. The scope of an individual robot's autonomy depends on how sophisticated the robot's interaction with its environment has been designed to be.

[12] This criterion distinguishes autonomous robots in the sense used here from software agents. On the latter and their treatment in private law, see *Beck*, AI & Society 31 (2016), 473, 478. See further on the embodiment criterion in robotics, euRobotics, The European Robotics Coordination Action, Suggestion for a green paper on

surroundings and able to act upon its surroundings, and, most importantly, it is equipped with artificial intelligence (in the following: AI). AI means that the robot has the ability to independently take decisions and implement them in the outside world.[13] Hence, the robot is neither exclusively operated remotely by a human being, nor does it not only execute preliminarily programmed or trained functions. Instead, it is self-learning. As such, it is able to learn new behaviors and reactions through the interaction with its environment, anticipate certain situations, and adapt its behavior to its environment and changes in its surroundings.

Such autonomous robots are able to perform activities that used to be exclusively human. This is the technology's great innovation. Yet, these characteristics also have ambivalent consequences when it comes to criminal liability. First, for the average observer/user who has no detailed knowledge of the data generated by the AI and the algorithms governing the AI's evaluation of generated data, it is neither entirely predictable nor controllable how an autonomous robot will act in a certain situation.[14] Second, the cause of a robot's self-learnt act or omission cannot be immediately traced back to a specific action of a human actor, such as the robot's manufacturer, distributor, owner, or user.[15] Certainly, it is a human who manufactures or uses the robot and as such creates the possibility for the robot to act in the first place. But the robot's autonomous act itself is neither immediately induced nor mastered by a human. These characteristics distinguish autonomous robots and their influence on criminal law from progresses in technology whose introduction we have witnessed in the past. To be clear: Technological revolutions have always posed challenges for conventional doctrines of criminal liability due to the unpredictability of new risks and courses of causation and the lack of established moral, social, and consequently legal standards of care in the context of the development, production, distribution, and usage of new technologies.[16] Yet, the interference by an artificially intelligent protagonist who acts independently from human determination and control adds an unprecedented dimension. It renders the application of traditional rules of criminal liability and the recourse to a natural person behind the robot even more difficult.

III. Case study

Nowadays, robots with differing degrees of autonomous capacities are used in multiple contexts. Practical examples range from intelligent drones supporting military operations to robots involved in the care for the elderly or medical surgery. Yet, the field that symbolizes

issues in robotics, Contribution to Deliverable D3.2.1. on ELS issues in robotics, p. 60 seq. (Christophe Leroux, Roberto Labruto ed., 31.12.2012).

[13] Cf., e.g., European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), P8_TA(2017)0051, para. AA. For different degrees of autonomy in the context of robots, see also *Hildebrandt*, Criminal Liability and "Smart" Environments, in: Duff/ Green, Philosophical Foundations of Criminal Law (2011), p. 507, 514. The degree of autonomy referred to here resembles most closely *Hildebrandt's* third category of "smart" devices.

[14] If the AI itself determined the algorithms governing its evaluation of generated data, nobody could ever foresee its actions. However, IT-technology is not that advanced yet, and it is currently unclear whether it will ever be.

[15] Note, though, that usually only some of an autonomous robot's behavior is self-learnt, while other behavior is determined by preset programming or user-influence. Hence, only some of the robot's behavior can no longer be traced back to the concrete actions (and mistakes) of a specific human being.

[16] For more details, see *Beck*, Robotics and Autonomous Systems 86 (2016), 138, 139 seq.

digitalization of society and future use of autonomous robots most graphically is the technology of autonomous driving. Issues of criminal responsibility for harm caused by an autonomous robot can therefore best be illustrated by an example drawn from this field.

Therefore, let us imagine the following futuristic, but not implausible scenario occurring on a future airport privately operated by person A:[17] On this airport, airplanes do not stop directly at the terminal, but somewhere on the maneuvering field. Passengers exit the plane on the maneuvering field and a bus transports them to the terminal. The bus is not navigated by a human driver, but by an artificially intelligent system. The AI enables the bus to determine by itself which route to take on the maneuvering field. Through interaction with its environment and the observation of other vehicles on the maneuvering field, the bus constantly self-learns and optimizes the route it takes. On the airport ground is a construction site. While the bus navigates around the site, it repeatedly observes that other vehicles – which are specifically equipped to move on rough terrain, e.g. with chain wheels – cross the construction site. The AI infers from this observation that vehicles can cross the construction site. One day, not realizing the different equipment of the bus and the construction vehicles, the AI independently decides to navigate the bus across the construction site because it identifies this as the overall most efficient route. The rough terrain is not suitable for the bus, though. Consequently, the bus overturns and a passenger is badly injured.

Let us further assume that the manufacturer and programmer of the autonomous vehicle had ensured that the vehicle had met all current scientific and technical standards and its safety for customers had been adequately tested before marketing, and that they had continuously monitored the vehicle afterwards. The installed safety mechanisms would have normally detected the rough terrain and its unsuitability for the bus, stopping it from taking the dangerous route. However, because of the self-learnt information, the AI unforeseeably de-prioritized the information given by the safety mechanism and took the fatal route.

If we leave aside the manufacturer's and programmer's potential criminal liability,[18] the question that immediately arises is: can person A who operates the airport and uses the autonomous bus for her purposes be held criminally liable for the passenger's injury?[19]

[17] Under German public law, fully autonomous cars are not yet allowed to drive on public streets. The related public law issues do not bear on the case-study, though, since it is designed to occur on private ground. – Should the legislator legalize autonomous cars on German streets in the future, this would also affect the determination of the standard of care in the context of negligence liability: The mere use of such cars could then be considered as a "permitted risk" (erlaubtes Risiko) what would exclude criminal liability.

[18] It is controversial whether the traditional principles of liability for unsafe products, which are used to determine the criminal responsibility for negligence of producers of unsafe products, apply when an autonomous robot has caused harm. Some authors doubt that such a robot could be classified as a defective product, see, e.g., *Beck*, AI & Society 31 (2016), 473, 475. Others argue that liability for harm caused by an autonomously acting robot is governed by the same rules that apply in other product liability cases, see, e.g., *Gless/Silverman/ Weigend*, New Crim. L. Rev. 19 (2016), 412, 427 seq. *Gless/Silverman/Weigend*, however, promote relaxing the liability standards for producers of autonomous agents *de lege ferrenda*, however, *id.* 430 seq.

[19] In order to focus on issues idiosyncratic to robotics, the case study assumes that A is not a legal, but a natural person, thereby excluding issues of corporate criminal liability.

IV. Criminal liability of the human actor behind the robot[20]

Under German criminal law, A would be held criminally liable for negligence if causation of the passenger's injury could be attributed to her behavior, if she could have foreseen the eventual harm and course of causation, and if she had failed to exercise the due care necessary to avert the foreseeable harm.[21] The involvement of an autonomous robot affects all these requirements.

1. Causal behavior

It might already be questioned which behavior of A could at all give rise to criminal liability. In the immediately harmful situation – when the artificially intelligent system decided to cross the rough terrain, causing the bus to overturn and consequently injure the passenger – A did not act. Instead, in this situation, the autonomous robot was the sole decider and actor. Therefore, to hold A criminally liable, we have to refer to a behavior much earlier in time; i.e., A's decision to use the autonomous robot to transport passengers. This decision set the first causal condition for a number of incidents that finally resulted in the passenger's injury. Hence, the only action that we could blame A for is that he used the autonomous agent in the first place. One must not overlook, though, that this premise has potentially far reaching impacts: If the mere usage of an autonomous agent triggered criminal liability, the digital revolution on an entire technology might be impeded before it has even really started.

2. Foreseeability of harm

The second condition of liability is that A, at the time that he decided to use the robot, must have been able to foresee the eventual harm as well as the course of causation. Here, we encounter the next challenge particular to the field of autonomous robots. As highlighted already, it is one of the very characteristics of autonomous robots that they take their decisions independently and that their behavior, consequently, cannot be foreseen in detail. When A decided to use the artificially intelligent bus, he knew that it would independently analyze the information that it acquired from its environment and would act autonomously in response to the results of its analysis. Generally, this autonomy and the benefits associated with it would most likely be the very incentives for employing autonomous agents. However, it is also this autonomy that renders the bus – to some extent – inherently dangerous.

These circumstances suggest two mutually exclusive conclusions. Either, A could aver that because the robot acted autonomously, it was impossible for A to foresee the robot's particular behavior and consequently the events resulting in the harmful accident. Consequently, A would be released from any liability for the robot's harmful actions. Inversely, it could be asserted that precisely because the robot acted independently, A had to expect "anything". The latter approach would result in the assumption that A could foresee any and all harm caused by the robot.

[20] The following discussion only regards negligent behavior. Judgments differ, of course, when a human being intentionally or knowingly uses an autonomous robot as a tool to cause harm to others.

[21] *Wessels/Beulke/Satzger*, Strafrecht Allgemeiner Teil, 47th ed. 2017, para. 935.

Under traditional German criminal law doctrine, it is considered sufficient if the actor can foresee the eventual harm and the general chain of causation.[22] Only a course of causation that is so unusual that nobody could reasonably anticipate it is considered unforeseeable. Since A knew the general peril inherent in the employment of an autonomous vehicle, the traditional doctrine would most likely be interpreted as to imply that the chain of causation between the autonomous agent's act and the eventual harm was not so unusual as to exclude foreseeability.[23] The underlying rationale would be that if A decided intentionally to employ an autonomous robot whose actions are generally unpredictable, she could not legitimately assert that this very unpredictability bared the foreseeability of harm and thus relieved her from liability.[24] The consequence of this line of argument, though, is that the foreseeability of harm requirement in traditional negligence doctrine would no longer function as a restriction on liability in the context of autonomous robots.

3. Violation of due care

Furthermore, A must have failed to exercise the due care necessary to avert the foreseeable harm. The requirements of due care are usually derived from a reasonable-person-standard.[25] Decisive is what a reasonable person would have *ex ante* considered necessary in order to avoid damages in a comparable situation. The first challenge in the context of autonomous robots in this regard is that we lack experiences with the handling of this evolving technology. It is difficult to define how a reasonable actor would handle a future complex technology that does not resemble anything that we know so far.[26]

Additionally, the reasonable person's behavior is often determined with reference to the foreseeability of harm.[27] It is argued that a reasonable person would refrain from undertaking a behavior if she could foresee that this behavior could jeopardize or harm a legally protected interest. Hence, acting despite the foreseeability of potential harm is regarded as violating due care. Since the behavior of an autonomous agent is predictably unpredictable, and consequently, the risk that it harms a legally protected interest can never be completely eliminated, under this theory, already the mere use of an autonomous agent would have to be qualified as violating due care. This reasoning has been criticized as not meeting the realities and demands of a modern digitalized risk society, however. The mere foreseeability of potential harm, it is argued, could not be used as the premier criterion for the determination of due care in the context of new technologies because this would essentially lead to the

[22] Cf. *Sternberg-Lieben/Schuster*, in: Schönke/Schröder, Strafgesetzbuch, 29th ed. 2014, § 15 para. 180; *Kühl*, in: Lackner/Kühl, Strafgesetzbuch, 28th ed. 2014, § 15 Rn. 46.

[23] See on this issue generally *Beck*, Robotics and Autonomous Systems 86 (2016), 138, 139.

[24] Concluding the same *Gleß/Weigend*, ZStW 126 (2014), 561, 581 seq.; *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 426 seq. Ambivalent *Beck*, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?, in: Beck/Meier/Momsen, Cybercrime and Cyberinvestigations (2015), p. 9, 26, who promotes an adjustment of the foreseeability requirement in the context of robotics *de lege ferenda*.

[25] *Kühl*, in: Lackner/Kühl, Strafgesetzbuch, 28th ed. 2014, § 15 para. 37; *Wessels/Beulke/Satzger*, Strafrecht Allgemeiner Teil, 47th ed. 2017, para. 943.

[26] *Beck*, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?, in: Beck/Meier/Momsen, Cybercrime and Cyberinvestigations (2015), p. 9, 25.

[27] *Wessels/Beulke/Satzger*, Strafrecht Allgemeiner Teil, 47th ed. 2017, para. 942.

prohibition of many new advantageous technologies.[28] Despite this criticism, the standard of care that courts would most likely - currently - apply is that a reasonable person would be expected to observe and control an autonomous robot such intensively that she could stop and avoid any potentially harmful robot action.[29] The final consequence of this approach is that a person who employs an autonomous robot would be held responsible for all damages caused by the robot.[30] Hence, in the case study, A would be blamed for not having exercised due care because she employed the robot-bus without stopping it from making the fatal navigation decision.

4. Attribution of harm

One might consider challenging this conclusion with the argument that the robot's autonomous behavior precludes the attribution of harm caused by the robot to the causal behavior of the human behind the robot. Along these lines, traditional German criminal law theory recognizes that the autonomous interference by another person with a causal chain of events can limit the attribution of harm to the causal act of a merely negligent first actor.[31] The analogous application of this theory to an autonomous robot's self-learnt intervention would reflect the very fact that autonomous robots, when acting on the basis of their AI, act independently. Eventually, it would result in equating an autonomous robot's intervention with a human being's intervention under criminal law theory. Several arguments weight against this analogy, however.[32] Most importantly, the analogy would result in a large liability vacuum. If two humans set intervening causes resulting in harm and one is discharged, the other can still be held to account. If a human and an autonomous robot set intervening causes resulting in harm and the human is discharged, at least under current law, no-one could be held criminally responsible.[33] Finally, if one intended to apply the analogy nonetheless, its application would at least need to be restricted to robot's acts that are driven by its AI and can as such be truly attributed to the robot itself.

[28] *Beck*, Robotics and Autonomous Systems 86 (2016), 138, 141; in this direction also *Gleß/ Weigend*, ZStW 126 (2014), 561, 582 seq.

[29] Cf. District Court Munich NJW-RR 2008, 40 (note though, that this judgment deals with negligence liability under civil law); in more depth *Thommen/Matjaz*, Die Fahrlässigkeit im Zeitalter autonomer Fahrzeuge, in: Jositsch/Schwarzenegger/Wohlers, Festschrift für Andreas Donatsch (2017), p. 273, 287 seq.; differently, though, at least in the context of autonomous cars *Lutz*, NJW 2015, 119, 121. – A person employing an autonomous robot could neither defend herself under the principle of reliance ("Vertrauensgrundsatz") with the argument that she expected the robot to "act" carefully and not to cause harm to others, see in more depth *Beck*, Robotics and Autonomous Systems 86 (2016), 138, 141.

[30] An exception could apply to damages caused by a construction defect of the robot. For such damages, the manufacturer could be held liable according to the standards of negligence liability for defective products, see *supra* note 16.

[31] See, *Eisele*, in: Schönke/Schröder, Strafgesetzbuch, 29th ed. 2014, Vor § 13 para. 100 seq.; *Sternberg-Lieben/Schuster*, in: *id.*, § 15 para. 171; *Wessels/Beulke/Satzger*, Strafrecht Allgemeiner Teil, 47th ed. 2017, para. 968 seq.

[32] See, e.g., *Beck*, Robotics and Autonomous Systems 86 (2016), 138, 141; *Gleß/Weigend*, ZStW 126 (2014), 561, 588; *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 432.

[33] Regarding a criminal responsibility of autonomous robots see *sotto* section V. If autonomous robots were held criminally liable, the result of the analysis may be different.

5. The way ahead: reconciling competing interests by shifting risks

The previous analysis has demonstrated that under current law, a person who employs an autonomous robot faces great risks of criminal liability. Also A, the operator of the fictitious future airport, would be held criminally liable for the passenger's injury caused by the robot-bus. The question is: does this legal standard serve the interests of society?

Autonomous robots, like the robot-bus in the case study, will be employed for the very reason that they can independently assume tasks. The application of a standard of care that requires the user to constantly observe and control the autonomous robot in order to avoid a potentially harmful action by the robot would render the delegation of a task to an autonomous robot pointless. Eventually, the technology's purpose and benefit would be eliminated. What is more: Such a great risk of criminal liability would most likely impede the use of autonomous robots altogether. That is because criminal liability is personal. Other than the risk of civil liability, a user cannot mediate the risk of criminal liability, e.g., by concluding an insurance contract.

This result appears ambivalent if we take into account that autonomous robots can also provide multiple benefits to individuals as well as to society. One may just consider autonomous cars: they may unleash resources, equalize access to mobility of disadvantaged groups, enhance traffic efficiency and most importantly – compared to human drivers – improve traffic safety.[34] If society regards such advantages desirable, it makes little sense to establish such a strong disincentive to the use of autonomous robots. Moreover, from the perspective of justice, it appears unfair if society on the one hand welcomes the benefits associated with the use of a new technology, and on the other hand shifts all risks that come with its use onto the individual by holding him criminally liable if something goes wrong. Finally, such a broad criminal liability would immensely restrict the personal autonomy of citizens to decide whether or not to employ autonomous robots.

Therefore, society needs to enter into a dialogue about a balanced approach to criminal liability in the context of autonomous robots. It has to decide to which extent it is willing to tolerate risks that are necessarily associated with the use of autonomous robots in order to benefit from the advantages generated by this technology. To this extent, society needs to shift the risk that an autonomous robot can cause harm from the individual user onto society at large by limiting individual criminal liability accordingly.[35] The legislature, in cooperation with the legal profession and academia, then has to appropriately reflect this societal decision by designing standards of reasonable care that are specifically tailored to the employment of

[34] The hope that autonomous cars are safer than human drivers has not even be severely challenged by the sadly famous case of a fatal accident involving a new Tesla car with level two autonomy that occurred in the US in 2016. The US National Transportation Safety Board detected that the primary factors contributing to the accident had to be ascribed to human errors, not to technological deficiency. The Board, therefore, concluded that autonomous cars, despite not being perfect, are still safer than human divers, see *Tristan Greene*, US authorities conclude last year's fatal Tesla crash was mostly human error, the next web: artificial intelligence, published Sept. 12, 2017, https://thenextweb.com/artificial-intelligence/2017/09/12/tesla-doesnt-deserve-all-the-blame-in-fatal-2016-crash/ (last accessed Oct. 23, 2017).

[35] Similar opinions share, e.g., *Beck*, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?, in: Beck/Meier/Momsen, Cybercrime and Cyberinvestigations (2015), p. 9, 30 seq.; *Thommen/Matjaz*, Die Fahrlässigkeit im Zeitalter autonomer Fahrzeuge, in: Jositsch/ Schwarzenegger/Wohlers, Festschrift für Andreas Donatsch (2017), p. 273, 289 seq.; *Gleß/Weigend*, ZStW 126 (2014), 561, 583 seq.; *Gless/Silvermann/Weigend*, New Crim. L. Rev. 19 (2016), 412, 430 seq.

autonomous robots and reconcile the competing interests. Technically, this could be achieved by defining certain risks associated with the use of AI as "admissible" what would in turn lower the required standard of care.[36] A person harmed by a robot would consequently be regarded as a victim of a socially accepted risk, not as a victim of the negligent wrongdoing of any particular person. In order to allow victims of autonomous robots to nevertheless seek redress for the harm suffered by a robot's act, the introduction of instruments other than criminal liability should be considered, too, such as the creation of a compulsory insurance system for robot users or a victim fund. Finally, to achieve an appropriate balance between the interests involved, it could further be contemplated to adjust the reduction of individual criminal liability according to characteristics of certain types of robots, such as whether a specific robot type is particularly useful for society, or whether the intended use of a robot is more or less likely to jeopardize the interests of innocent bystanders.

V. Criminal liability of autonomous robots

Admittedly, more liberal standards of care regarding the handling of autonomous robots could result in situations in which an autonomous robot would harm a third party and no human would be held criminally liable. Furthermore, already today, attempts to hold individuals criminally liable for damages caused by robots often fail. The main reason is that causation cannot be proven beyond a reasonable doubt.[37] *Ex post*, it is often almost impossible to determine whether the robot malfunctioned because of a machine defect for which the manufacturer could be held to account, a user's error or training of the robot for which the user could be held liable, or whether the machine's harmful action is a result of its autonomous self-learning. Hence, in many instances, all human actors can escape criminal responsibility.

1. Recent proposals to hold autonomous robots liable

In order to address this liability vacuum, a rather futuristic solution has been suggested, i.e., holding autonomous robots themselves liable for their acts or omissions that damage third parties.[38] Most prominently, the European Parliament advocates in its resolution to the Commission on a civil law for robots the creation of a new legal category, specifically designed for the most sophisticated autonomous robots: an electronic personhood.[39] The electronic person is envisioned as a subject of specific rights and obligations, including personal liability for damaging third parties. The Parliament's suggestion was inspired by proposals of the Robolaw project, a study funded under the European Commission's 7th

[36] For alternative dogmatic solutions see *Gleß/Weigend*, ZStW 126 (2014), 561, 584 footnote 93.

[37] See, e.g., *Beck*, JR 2009, 225, 226 seq.; *Beck*, Roboter und Cyborgs – erobern sie unsere Welt?, in: Beck, Jenseits von Mensch und Maschine, Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 9, 15; particularly on (partly) autonomous cars *Lutz*, NJW 2015, 119, 120.

[38] E.g., *Beck*, Technisierung des Menschen – Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept „Verantwortung", in: Gruber/Bung/Ziemann, Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft (2014), p. 173, 181 seq.; *Schuhr*, Neudefinition tradierter Begriffe, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 13, 18 seq. See also the following references.

[39] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), P8_TA(2017)0051, number 59 (f.). Note though, that the Parliament expressly makes this a proposal for some time in the future, *id.* number 56.

Framework Program for Research and Technological Development.[40] The idea to create a legal category of electronic personhood was controversial during the resolution's drafting process, however. In preparation of the resolution, the European Parliament's Committee on Legal Affairs had requested a study on European Civil Law Rules on Robotics which was commissioned, supervised and published by the Policy Department for Citizens' Rights and Constitutional Affairs. This study strongly refused the vision to acknowledge the figure of an electronic person as civilly liable legal entity, calling it "fanciful visions about robots [...] based on science fiction"[41]. Nevertheless, the proposal to acknowledge electronic personhood is included in the final Parliament resolution.

The proposal to create the category of electronic personhood is primarily inspired by the motive to establish an effective system of liability. The electronic person shall bundle the legal responsibilities of the human beings behind the machine (e.g., manufacturer, seller, user, etc.). They shall be responsible for stocking the electronic person with financial assets.[42] As such, the electronic person would function as "a 'tangible symbol' for the cooperation of all the people creating and using that specific robot".[43] A wronged third party would no longer need to prove liability of any individual human behind the robot, but could directly sue the electronic person. Advantages of such a legal figure are said to include the prevention of liability gaps, a fair distribution of liability among the humans behind the robot, and an alleviation of proof for the victim. Looked at it more closely, the suggestion of an electronic person resembles quite some features and objectives of the existing category of legal persons.[44]

So far, proponents of electronic personhood only campaign for civil liability. The debate on the liability of legal persons suggests, though, that once civil liability might have been established, *criminal* responsibility of future electronic persons would most likely soon be called for, too. Compared to any criminal responsibility of legal persons, a potential criminal responsibility of future electronic persons might even suggest itself more. That is because (most) electronic persons would possess a physical structure making them more visible and able to interact in the real world context than the abstract figure of a legal person.[45] Furthermore, an autonomous robot can – to some extent – take decisions independently of any

[40] See euRobotics, The European Robotics Coordination Action, Suggestion for a green paper on issues in robotics, Contribution to Deliverable D3.2.1. on ELS issues in robotics, p. 60 seq. (Leroux/Labruto ed., Dec. 31, 2012). This project intended to assess whether existing EU regulations are sufficient to address the various legal problems posed by robotics technology, and to ensure that EU law provides sufficient conditions to incentivize European innovation in the robotics sector.

[41] Policy Department C: Citizens' Rights and Constitutional Affairs, European Civil Law Rules on Robotics - Study for the JURI Committee, October 2016, PE 571.379, *available at* http://www.europarl.europa.eu/committees/fr/supporting-analyses-search.html (last accessed Oct. 23, 2017), p. 5, 14 ff.

[42] *Beck*, AI & Society 31 (2016), 473, 480.

[43] euRobotics, The European Robotics Coordination Action, Suggestion for a green paper on issues in robotics, Contribution to Deliverable D3.2.1. on ELS issues in robotics, p. 61 (Leroux/Labruto ed., Dec. 31, 2012); similarly *Beck*, AI & Society 31 (2016), 473, 479.

[44] Proponents of the idea even develop characteristics of the electronic person based on characteristics of legal persons, see euRobotics, The European Robotics Coordination Action, Suggestion for a green paper on issues in robotics, Contribution to Deliverable D3.2.1. on ELS issues in robotics, p. 61 (Leroux/Labruto ed., Dec. 31, 2012); *Beck*, Technisierung des Menschen – Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept „Verantwortung", in: Gruber/Bung/Ziemann, Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft (2014), p. 173, 179, 181.

[45] Similar *Beck*, Robotics and Autonomous Systems, 86 (2016) 138, 142.

natural person whereas a legal person and its actions are eventually always restricted to the sum of the humans composing it. Despite this juxtaposition, one must not forget that German criminal law does not (yet) affirm criminal responsibility of legal persons and that the arguments for the refusal might be equally valid with regard to a potential criminal liability of a future electronic person, however.[46]

Finally, going beyond the calls for electronic personhood in private law, some scholars indeed already investigate options of a genuine criminal liability of autonomous robots.[47] Several of them confirm – partly under the condition that autonomous robots will once resemble human beings in characteristics such as intelligence, empathy and free moral agency – that autonomous robots could or even should be held criminally liable.[48]

2. Autonomous robots and personal autonomy

Criminal responsibility of autonomous robots could certainly ease releasing the human behind the robot from liability since it would create a new responsibility sphere preventing the otherwise evolving liability gap. This fact alone, however, cannot compensate the serious challenges posed by the suggestion to hold autonomous robots criminally liable under conventional criminal law theory. In this regard, it may, for example, be disputed whether autonomous robots are capable of "acting" in a way sufficient under criminal law theory, even if they possess an embodiment that allows them to physically impact their environment.[49] Most importantly, though, the question of criminal liability of autonomous robots brings us back to the matter of personal autonomy.

As highlighted above, criminal responsibility under German criminal law is theoretically and constitutionally founded on the concept of personal autonomy.[50] Only creatures can be held criminally liable who can be ascribed personal autonomy or, in the words of the German

[46] One should note, though, that Germany's position on criminal liability of legal persons is becoming the minority view in Europe where many countries have more or less recently introduced corporate criminal responsibility. For an overview of the relevant law in different countries, see *Göhler*, ZIS 2016, 219, 220 seq.; *Heine/Weißer*, in: Schönke/Schröder, Strafgesetzbuch, 29th ed. 2014, Vorb. §§ 25 seq. para. 124.

[47] E.g., *Beck*, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?, in: Beck/Meier/Momsen, Cybercrime and Cyberinvestigations (2015), p. 9, 27, 31; *Beck*, Technisierung des Menschen – Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept „Verantwortung", in: Gruber/Bung/Ziemann, Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft (2014), p. 173, 182; *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 415 seq., 435; *Gless/Weigend*, ZStW 126 (2014), 561, 566 seq., 570; *Hildebrandt*, Criminal Liability and "Smart" Environments, in: Duff/Green, Philosophical Foundations of Criminal Law (2011), p. 507, 526 seq., 532; *Hilgendorf*, Können Roboter schuldhaft handeln?, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 119 seq.; *Neuhäuser*, Roboter und moralische Verantwortung, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 269 seq.; *Neuhäuser*, Künstliche Intelligenz und ihr moralischer Standpunkt, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 23 seq.; *Schuhr*, Neudefinition tradierter Begriffe, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 13, 25 seq.; *Schuhr*, Willensfreiheit, Roboter und Auswahlaxiom, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 43 seq.

[48] See most of the authors named in *supra* note 45, except for *Neuhäuser*.

[49] See on this issue skeptical *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 419 seq.; *Gless/Weigend*, ZStW 126 (2014), 561, 571 seq.; affirmatively *Hilgendorf*, Können Roboter schuldhaft handeln?, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 119, 125 seq.

[50] See *supra* section II 1.

Federal Court, free and responsible moral self-determination. The reason is that personal autonomy is regarded a prerequisite of personal culpability, what is a condition for criminal responsibility and punishment. Hence, to hold autonomous robots criminally liable, under current doctrine, we would need to confirm that they possess personal autonomy in the sense that they have (1) the ability to distinguish between the right and the wrong, (2) the ability and freedom to decide in favor of the right and against the wrong, and (3) the ability to adapt their behavior according to this decision.

In this regard, it is important to note that robots acting according to programmed or trained algorithms clearly do not operate on such abilities. This is true, for instance, for a self-driving car that is programmed how to "behave" in a dilemmatic situation.[51] If such a car "decides", e.g., for the navigation option that causes the smaller amount of personal injuries, it does not do so based on its own moral evaluation of its options and its own decision in favor of the morally right, but it does so in execution of its pre-inscribed source code. Hence, even though it may look to the outside as if the self-driving robot-car took its own reflected moral decision, it does not.

The more challenging question is whether robots that act driven by their AI can be said to display personal autonomy. Some criminal law scholars affirm this, at least potentially for the future.[52] They argue that it is only a matter of time and research that autonomous robots will be able to engage in moral reasoning and acquire the necessary prerequisites or at least a functional equivalent of personal culpability.[53]

This reasoning has to be received with some reluctance. Contrary to first intuition, the primary problem, though, is not that personal culpability would necessarily be founded on a concept of "free will" as something sacrosanct human. Neuroscientific brain research suggesting the predetermination of human decision making has casted doubt upon the assumption that human beings possess the freedom to decide between different behavioral options based on their moral evaluations.[54] Despite this empirical research, the principle of guilt has not been dismissed in criminal law theory.[55] The reason is that the establishment of personal guilt is not considered an issue of empirically provable, neurological facts of free will, but the result of an ascription in social settings.[56] If criminal law is to complete its social function, the ascription of personal guilt must not be undertaken arbitrarily, however. Instead, it is required that the subject to whom personal guilt is attributed has the ability of moral self-reflection. An entity that is not capable of participating in a dialogue about and the formulation of moral standards and evaluating her own actions according to a moral reference system of right and wrong, can neither understand the ascription of personal culpability nor

[51] See on such occasions *supra* note 7.

[52] E.g., *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 423; *Gless/Weigend*, ZStW 126 (2014), 561, 579; *Hildebrandt*, Criminal Liability and "Smart" Environments, in: Duff/Green, Philosophical Foundations of Criminal Law (2011), p. 507, 532.

[53] *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 423.

[54] *Weißer*, GA 2013, 26 seq. with further references to relevant neuroscientific literature (p. 27).

[55] See particularly on the (ir)relevance of this brain research in the context of culpability of robots, *Gless/Weigend*, ZStW 126 (2014), 561, 574; *Hilgendorf*, Können Roboter schuldhaft handeln?, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 119, 129 seq.

[56] *Frister*, Strafrecht Allgemeiner Teil, 7th ed. 2015, p. 32 seq.; *Roxin*, Strafrecht Allgemeiner Teil, Band 1, 4th ed. 2006, § 19 para. 37; *Weißer*, GA 2013, 26, 36 seq.

the ethical reproval contained in criminal punishment. Holding such an entity criminally responsible would thus be senseless. Personal culpability of robots, therefore, is less a matter of whether autonomous robots could ever possess something opaque like a free will,[57] but whether they can be attributed the ability of moral self-reflection. Yet, also "merely" attributing an ability of moral self-reflection to robots is far from easy.

Authors affirming this ability argue that robots could be programmed with a system of "merits" and "demerits" for certain of their decisions, and that such a system could be treated as an analogue to human self-reflection on moral grounds.[58] This thesis, however, faces several problems. First, the formalization of ethical decisions, which would be necessary for a system that awards merit- and demerit-points to predefined factors, is extremely difficult.[59] Second, it is questionable whether the suggested programmed system of "merits" and "demerits" indeed resembles human self-reflection on moral grounds qualitatively in a way that would justify treating both equally. If a robot took a decision according to a programmed system of "merits" and "demerits", the robot would certainly calculate the value of different factors and react according to the result of this calculation.[60] However, the eventual result, i.e. the robot's "decision", would still be governed by a predetermined extrinsic value system. In other words, the robot would only figure out the programmer's *ex ante* abstractly defined preference for a given situation by calculating the merit- and demerit-points defined *ex ante* in its program code and would subsequently execute this humanely predetermined preference. Hence, despite the required calculating activities, the robot's "decision" would not be an autonomous decision based on moral considerations undertaken by the robot's AI, but a pre-determined action. In essence, it would be a decision of the human behind the robot responsible for the value system in the robot's source code. Furthermore, there is another important distinction between a robot's decision based on a programmed system of "merits" and "demerits" and a human decision based on an accomplished system of morality: Whereas the human can diverge from a learnt system of morality, the robot's programming would most likely not allow it to diverge from the result that it calculated to be the one in line with its initially inscribed program. Finally, the assumption of a robot acting based on a pre-programmed and inscribed value system presupposes that the robot would not itself participate in the societal dialogue about and the formulation of moral standards. This is worth noting because an entity's potential to participate in this interpretative communicative process, the community's recognition of an entity as participant thereto, as well as the entity's corresponding reflection of her recognition as valid participant are considered important factors for the establishment of an entity as a morally responsible actor.[61]

[57] Some authors, though, seek to develop theories that allow applying the concept of "free will" to robots, see *Schuhr*, Willensfreiheit, Roboter und Auswahlaxiom, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 43 seq.

[58] *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 423; *Gless/Weigend*, ZStW 126 (2014), 561, 576.

[59] *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 422 seq., recognize this difficulty. See regarding the (technical) challenges to design ethically responsible robots, *Weber*, Autonome und ferngesteuerte Kampfdrohnen, in: Gruber/Bung/Ziemann, Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft, 2nd ed. 2015, p. 267, 280 seq.

[60] Similarly, *Gless/Weigend*, ZStW 126 (2014), 561, 575.

[61] Cf. *Neuhäuser*, Künstliche Intelligenz und ihr moralischer Standpunkt, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs

Consequently, in order to create a digital correspondent that would be somewhat comparable to human self-reflection on moral grounds, at least two technical requirements would need to be met: Firstly, the robot would need to self-learn the system of "merits" and "demerits" with its AI and decide according to it. Only then, the moral decision could be considered as the robot's decision and not as the decision of the human behind it. The technical implementation of this requirement, however, is even more difficult than the design of the proposed programmed system of "merits" and "demerits" which is not even reality yet.[62] Secondly, the robot's source code would need to authorize the robot to deviate from a decision that it has calculated as being the morally correct decision. Whether this would be desirable, however, is an entirely different kettle of fish.

Finally, even if computer scientists succeeded in developing an artificially intelligent system that met the outlined technical requirements, it might still be arguable whether this would suffice for the ascription of moral self-reflection and consequently personal culpability. Doubts may remain because one might additionally require the robot to be capable of *empathetically* acknowledging the moral and social consequences of its actions (and of the penalty received therefor).[63]

As things stand today, even autonomous robots lack – and most likely will lack for some time to come – the necessary prerequisites of personal autonomy as a condition of personal culpability. Hence, under current doctrine, they cannot be held criminally responsible. That is why we have to ask us the initially raised question: Should we rethink personal autonomy as foundational requirement of criminal liability under German criminal law in order to tackle developments raised by robotics? Or in other words: Do we want a "new" criminal law for autonomous robots?

3. A "new" criminal law for autonomous robots?

Before one starts to seriously contemplate adapting criminal law theory so as to pave the way for a criminal responsibility of autonomous robots, one should reflect on the following critical remarks.

a) Constitutional guarantees

First and foremost, personal autonomy, or what the German Federal Court calls free and responsible moral self-determination is a prerequisite for the ascription of personal culpability, i.e., guilt. The guilt principle as a foundation of criminal law and punishment is

(2012), p. 23 seq.; *Neuhäuser*, Roboter und moralische Verantwortung, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 269, 276 seq.

[62] See regarding the issue whether a robot could adopt a legal or moral standpoint, *Neuhäuser*, Künstliche Intelligenz und ihr moralischer Standpunkt, in: Beck, Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs (2012), p. 23 seq.

[63] Cf. *Beck*, Brauchen wir ein Roboterrecht? Ausgewählte juristische Fragen zum Zusammenleben von Menschen und Robotern, in: Japanisch-Deutsches Zentrum, Mensch-Roboter-Interaktionen aus interkultureller Perspektive. Japan und Deutschland im Vergleich (2012), p. 124, 137 seq.; *Hildebrandt*, Criminal Liability and "Smart" Environments, in: Duff/Green, Philosophical Foundations of Criminal Law (2011), p. 507, 527 seq.; *Neuhäuser*, Roboter und moralische Verantwortung, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 269, 278 seq.

constitutionally guaranteed.[64] According to the German Federal Constitutional Court's judgment on the Lisbon treaty, the guilt principle together with the undergirding concept of personal autonomy is even enshrined in the supreme value of human dignity.[65] Hence, not even a constitutional amendment could impugn the validity of these principles. Against this constitutional background, the concept of personal autonomy could not simply be dispensed with in order to adapt criminal law theory so as to hold autonomous robots liable.

b) Issues of penal theory and punishment

Second, holding autonomous robots criminally responsible would also entail punishing them. Criminal punishment traditionally consists of two components: an ethical reproval of the culprit's behavior expressed by the state representing the public and a sanction.[66] The ethical reproval of a perpetrator's wrongful behavior mirrors the assumption that a perpetrator possesses personal autonomy and is hence able to understand the wrongfulness of her behavior and consequently the corresponding reproval. If robots are not conceived as agents endowed with the ability of moral self-determination and self-reflection, they can neither be deemed to apprehend and respond to an ethical reproach. Because of this interrelation, punishing robots that do not possess personal autonomy or a suitable equivalent would amount to a legal-moral farce.[67] The alternative – to dispense with the ethical reproval as distinguishing characteristic of criminal (as compared to, e.g., administrative) punishment – would eventually mean to dispense with criminal law as it exists today.

Furthermore, the system of criminal sanctions is geared toward human beings and difficult to adapt to nonhuman robots. As long as robots do not feature personal desires, such as a wish for freedom or own property, conventional sanctions, such as incarceration or fines, cannot have a comparable effect on robots as they have on humans.[68] One might consider inventing robot-specific sanctions, e.g., a software-reset or the destruction of the robot's physical structure. These tools might be appraised as approximating educative sanctions or even the death penalty for humans. The problem that remains, though, is whether robots as such are suitable recipients of preventive goals of punishment.[69] While this is already questionable as far as the goal of positive special prevention is concerned,[70] it is quasi inconceivable how the goal of positive general prevention could be met with regard to autonomous robots that lack personal autonomy. Moreover, if one was to share the emerging trend of evaluating penal sanctions from the perspective of the victim of a crime,[71] it appears arguable whether the victim would gain satisfaction by seeing the robot being punished; or phrased more

[64] German Federal Constitutional Court (Bundesverfassungsgericht) NJW 1977, 1525, 1532; NJW 2009, 2267, 2289.

[65] German Federal Constitutional Court (Bundesverfassungsgericht) NJW 2009, 2267, 2289.

[66] *Kühl*, in: Lackner/Kühl, Strafgesetzbuch, 28th ed. 2014, § 46 para. 1.

[67] Similarly, *Schuhr*, Neudefinition tradierter Begriffe, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 13, 26.

[68] Similarly, *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 424; *Gless/Weigend*, ZStW 126 (2014), 561, 577 seq.

[69] See on the different theories of punishment more generally *Joecks*, in: Münchener Kommentar zum StGB, 3rd ed. 2017, Einl. para. 52 seq.; *Kühl*, in: Lackner/Kühl, Strafgesetzbuch, 28th ed. 2014, § 46 para. 2 seq.

[70] For a critical assessment of this question, see *Neuhäuser*, Roboter und moralische Verantwortung, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 269, 283; more optimistic *Schuhr*, Neudefinition tradierter Begriffe, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 13, 26.

[71] See, e.g., *Hörnle*, Straftheorien, p. 37 seq. (2011).

graphically: would it help the injured passenger in our case study to see the robot-bus being transferred to a junk press in order to come to terms with the harmful experience?

Finally, robot-specific sanctions, such as a complete software-reset or the destruction of the robot's physical structure, could be unjustifiable from a rights' perspective.

c) Flipside of criminal responsibility: basic rights for robots?

The reference to the rights' perspective leads us to a third point: one condition of robots' criminal liability is that robots are obliged to obey criminal laws. An obligation to obey criminal laws, in turn, presupposes that robots enjoy – at least to a certain extent – freedom of action. Furthermore, the rule of law would require that robots are guaranteed procedural rights during their criminal prosecution. Hence, the flipside of a criminal responsibility of robots is that robots are granted at least certain basic rights. This observation raises a whole lot of questions, for instance, which concrete rights would have to be guaranteed: a right to life or to "digital existence", to liberty, to be treated equally to other robots during a criminal process or equally to other human defendants? One would also have to scrutinize how answers to these questions would affect the legitimacy of certain potential robot-specific sanctions. Moreover, under constitutional law and democratic theory, would robots then not also have to be included or at least also have the right to participate in the election of the legislature if that legislature defined criminal offenses encompassing also their behavior?[72] Beside these rather theoretical questions, the rights discourse could also have very practically implications. That is, if humankind was obliged to respect robots' rights, would it then still be able to "use" robots in the same manner as it employs machines today? If not, how would that consequence impact the robotics market?

No matter what the answers to these questions are, one thing should be clear: we cannot have the cake and eat it. Treating robots equally to humans with regard to criminal liability means that we would also have to treat them equally in other respects.

d) Digital idiosyncrasies of robots

Moreover, digital idiosyncrasies of robots impede their simple subsumption under a system of criminal responsibility meant and designed for humans.

From a scientific perspective, it is impossible for a robot today, and probably will remain so for a long time to come, to take part in general and legal life without a human pulling at least some of its strings. That is to say that only part of an autonomous robot's behavior can be attributed to its AI, whereas the rest is still determined by preset algorithms and user input. Therefore, when holding an autonomous robot criminally liable, we would have to guarantee that we hold it to account exclusively for its *autonomous* unlawful behavior. Otherwise we would violate the principle that one can only be held criminally responsible for own wrongful behavior. This requirement would lead to the phenotypic puzzling result that a robot would be called to account for some, but not all of its actions. Since it is not necessarily distinguishable from the outside which behavior is driven by AI and which behavior is caused by pre-set

[72] Cf. *Hildebrandt*, Criminal Liability and "Smart" Environments, in: Duff/Green, Philosophical Foundations of Criminal Law (2011), p. 507, 517.

programming or user influence, this result would be difficult to understand, and most likely also hard to accept for an average observer.

Furthermore, whereas a human being has a definite contour and a homogenous mind, the material structure of an autonomous robot is more difficult to define. The physical embodiment that we visually identify as “the robot” can have polymorph phenotypes and the technical infrastructure behind it can consist of multiple software components. The latter may interact, but can also exist and operate quite independently from each other. Further, while some of these components may be artificially intelligent, others may not. We would therefore have to decide what to do if just one out of several artificially intelligent components inside one physical robot made a wrongful decision. Would or even could we punish the entire physical structure including the hard- and software components that did not participate in the wrongful decision? Would it make a difference whether these “innocent” components are themselves equipped with AI? After all, if we declare artificially intelligent structures as criminal responsible, it might be unfair if one artificially intelligent component had to suffer from criminal punishment for an action that was mastered by another artificially intelligent component. Yet, if we only punished the one responsible component, e.g., by removing it from the physical robot-structure, we would need to assess how this would backfire on the robot as a whole.

These questions demonstrate that it might not be plausible to simply equate the figure of an “autonomous robot” with a human being and discuss criminal responsibility of an autonomous robot as if its physical embodiment was as predetermined as the natural body of a human and as if its different software components were the simple equivalent of an homogenous human brain. This is not to say that we could not solve these questions. However, we should be cautious not to over-simplify comparisons.

e) Ultima ratio of criminal law

Finally, the application of criminal law must always be governed by the *ultima ratio* principle. This holds especially true in the context of new risky technologies. The ability of criminal law to control behavior in modern digitalized risk societies is particularly limited.[73] Often, in this context, criminal law is already (miss)used for mere symbolic purposes. Against this background, it remains arguable whether extending criminal liability to a new entity would be wise. Instead, other instruments might be better suited to channel and reconcile the advantages and perils associated with the emerging technology of robotics than criminal law.

Such alternatives could, for instance, include compulsory insurance systems, licensing requirements or best practice standards for the development and use of robots, funds for indemnifying those who have suffered harm from a robot’s action, to a limited extend strict civil liability of the human beings behind the robot, and maybe even a public authority monitoring the development and behavior of autonomous robots.[74] These alternatives need to

[73] See on so called risk criminal law („Risikostrafrecht“), e.g., *Prittwitz*, „Feindstrafrecht“ als Konsequenz des „Risikostrafrechts“, in: Vormbaum, Kritik des Feindstrafrechts (2009), p. 169 seq.

[74] Other authors further propose penal prohibitions of certain autonomous robots or of their employment for certain purposes, e.g., *Beck*, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?, in: Beck/Meier/Momsen, Cybercrime and Cyberinvestigations (2015), p.

be discussed and weighted in order to establish a balanced system for the handling of autonomous robots. Then, such alternatives could reasonably close the liability gap that might emerge if we simultaneously lowered criminal liability of the user of autonomous robots and decided not hold the robots themselves criminally liable.

VI. Conclusion

The emerging technology of autonomous robots touches our perception of the foundations of criminal liability in an unprecedented way. That is because autonomous robots will take independent decisions. Nevertheless, currently, it is neither necessary nor desirable to redefine personal autonomy as a foundational requirement of criminal liability. At least as technical developments stand today and likely will remain for some time to come, holding autonomous robots criminally liable is not a feasible solution to rising challenges posed by digitalization. Instead, alternative measures in civil and public law should be considered in order to close a potentially emerging liability gap. With regard to the liability of the humans behind the robot, society has to enter into a debate about how much risk it is willing to accept in return for the benefits associated with the use of AI. Legal academia and politics then have to reflect this debate by defining legal standards of care that are specifically tailored to the handling of autonomous robots.

The European Union has a common market and open boarders. Autonomous robots will therefore be developed by international research groups and exported, sold, and most importantly used internationally. Questions of criminal liability in the context of autonomous robots, therefore, have an immediate trans-boarder link. Consequently, the debate about these issues must take place internationally to ensure solutions that are suitable for a globalized world, foster further research on new technologies in the benefit of society, and address the people's fear with regard to this new technology.[75] Comparative law symposia offer the ideal context for this necessary debate.

VII. Bibliography

Susanne Beck, Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood, Robotics and Autonomous Systems 86 (2016), 138 seq.

Susanne Beck, The problem of ascribing legal responsibility in the case of robotics, AI & Society 31 (2016), 473 seq.

Susanne Beck, Technisierung des Menschen – Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept „Verantwortung", in: Malte-Christian Gruber, Jochen Bung, Sascha Ziemann (editors), Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft, 2nd ed. 2015, p. 173 seq.

9, 31 seq.; *Neuhäuser*, Roboter und moralische Verantwortung, in: Hilgendorf, Robotik im Kontext von Recht und Moral (2014), p. 269, 284 seq.

[75] Differently *Gless/Silverman/Weigend*, New Crim. L. Rev. 19 (2016), 412, 436 who claim that each society needed to answer for itself the question of criminal responsibility in the context of the development and operation of autonomous robots.

Susanne Beck, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?, in: Susanne Beck, Bernd-Dieter Meier, Carsten Momsen (editors), Cybercrime and Cyberinvestigations, 2015, p. 9 seq.

Susanne Beck, Roboter und Cyborgs – erobern sie unsere Welt?, in: Susanne Beck (editor), Jenseits von Mensch und Maschine, Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, p. 9 seq.

Susanne Beck, Brauchen wir ein Roboterrecht? Ausgewählte juristische Fragen zum Zusammenleben von Menschen und Robotern, in: Japanisch-Deutsches Zentrum (editor), Mensch-Roboter-Interaktionen aus interkultureller Perspektive. Japan und Deutschland im Vergleich, 2012, p. 124 seq.

Susanne Beck, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 2009, 225 seq.

Beatrice Brunhöber, Individuelle Autonomie und Technik im Körper, in: Susanne Beck (editor), Jenseits von Mensch und Maschine, Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, p. 77 seq.

euRobotics, The European Robotics Coordination Action, Suggestion for a green paper on issues in robotics, Contribution to Deliverable D3.2.1. on ELS issues in robotics (Christophe Leroux, Roberto Labruto editor, Dec. 31, 2012)

Helmut Frister, Strafrecht Allgemeiner Teil, 7th edition, 2015

Sabine Gless/Emily Silverman/Thomas Weigend, If robots cause harm, who is to blame? Self-driving cars and criminal liability, New Crim. L. Rev. 19 (2016) 412 seq.

Sabine Gleß/Thomas Weigend, Intelligente Agenten und das Strafrecht, ZStW 126 (2014), 561 seq.

Johanna Göhler, Tagungsbericht: Bewältigung der Finanzkrise mit den Mitteln des Strafrechts - Europäische Standards für die Voraussetzungen individueller und kollektiver strafrechtlicher Haftung?, ZIS 2016, 219 seq.

Erik Hilgendorf, Können Roboter schuldhaft handeln?, in: Susanne Beck (editor), Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, p. 119 seq.

Tatjana Hörnle, Straftheorien, 1st ed. 2011

Lackner/Kühl, Strafgesetzbuch Kommentar, 28th edition 2014

Lennart S. Lutz, Autonome Fahrzeuge als rechtliche Herausforderung, NJW 2015, 119 seq.

Mireille Hildebrandt, Criminal Liability and "Smart" Environments, in: R. A. Duff, Stuart P. Green (editors), Philosophical Foundations of Criminal Law, 2011, p. 507 seq.

Münchener Kommentar zum Strafgesetzbuch, 3rd edition 2017 (Wolfgang Joecks, Klaus Miebach editors)

Christian Neuhäuser, Roboter und moralische Verantwortung, in: Eric Hilgendorf (editor), Robotik im Kontext von Recht und Moral (2014), p. 269 seq.

Christian Neuhäuser, Künstliche Intelligenz und ihr moralischer Standpunkt, in: Susanne Beck (editor), Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, p. 23 seq.

Cornelius Prittwitz, „Feindstrafrecht" als Konsequenz des „Risikostrafrechts", in: Thomas Vormbaum (editor), Kritik des Feindstrafrechts, 2009, p. 169 seq.

Claus Roxin, Strafrecht Allgemeiner Teil, Band 1: Grundlagen – Der Aufbau der Verbrechenslehre, 4th edition 2006

Jan Schuhr, Willensfreiheit, Roboter und Auswahlaxiom, in: Susanne Beck (editor), Jenseits von Mensch und Maschine. Ethische und Rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, p. 43 seq.

Jan Schuhr, Neudefinition tradierter Begriffe, in: Eric Hilgendorf (editor), Robotik im Kontext von Recht und Moral, 2014, p. 13 seq.

Schönke/Schröder, Strafgesetzbuch Kommentar, 29th ed. 2014

Marc Thommen/Sophie Matjaz, Die Fahrlässigkeit im Zeitalter autonomer Fahrzeuge, in: Daniel Jositsch, Christian Schwarzenegger, Wolfgang Wohlers (editors), Festschrift für Andreas Donatsch, 2017, p. 273 seq.

Jutta Weber, Autonome und ferngesteuerte Kampfdrohnen. Über „Revolution in Military Affairs" und den Traum vom automatisierten Krieg, in: Malte-Christian Gruber, Jochen Bung, Sascha Ziemann (editors), Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft, 2nd ed. 2015, p. 267 seq.

Philipp Weber, Dilemmasituationen beim autonomen Fahren, NZV 2016, 249 seq.

Thomas Weigend, Notstandsrecht für selbstfahrende Autos?, ZIS 2017, 599 seq.

Bettina Weißer, Ist das Konzept strafrechtlicher Schuld nach § 20 StGB durch die Erkenntnisse der Neurowissenschaften widerlegt?, GA 2013, 26 seq.

Johannes Wessels/ Werner Beulke/ Helmut Satzger, Strafrecht Allgemeiner Teil. Die Straftat und ihr Aufbau, 47th edition 2017