If Robots Cause Harm, Who Is to Blame? Self-Driving Cars and Criminal Liability

Abstract:	The fact that robots, especially self-driving cars, have become part of our daily lives raises novel issues in criminal law. Robots can malfunction and cause serious harm. But as things stand today, they are not suitable recipients of criminal punishment, mainly because they cannot conceive of themselves as morally responsible agents and because they cannot understand the concept of retributive punishment. Humans who produce, program, market and employ robots are subject to criminal liability for intentional crime if they knowingly use a robot to cause harm to others. A person who allows a self-teaching robot to interact with humans can foresee that the robot might get out of control and cause harm. This fact alone may give rise to negligence liability. In light of the overall social benefits associated with the use of many of today's robots, however, the authors argue in favor of limiting the criminal liability of operators to situations where they neglect to undertake reasonable measures to control the risks emanating from robots.

SCHOLARONE™
Manuscripts

If Robots Cause Harm – Who Is to Blame?

Self-Driving Cars and Criminal Liability

Abstract

The fact that robots, especially self-driving cars, have become part of our daily lives raises novel issues in criminal law. Robots can malfunction and cause serious harm. But as things stand today, they are not suitable recipients of criminal punishment, mainly because they cannot conceive of themselves as morally responsible agents and because they cannot understand the concept of retributive punishment. Humans who produce, program, market and employ robots are subject to criminal liability for intentional crime if they knowingly use a robot to cause harm to others. A person who allows a self-teaching robot to interact with humans can foresee that the robot might get out of control and cause harm. This fact alone may give rise to negligence liability. In light of the overall social benefits associated with the use of many of today's robots, however, the authors argue in favor of limiting the criminal liability of operators to situations where they neglect to undertake reasonable measures to control the risks emanating from robots.

I. Introduction

Robots have arrived in our daily lives. Whereas in the past, images of autonomous vehicles were the stuff of screenwriters and science fiction novelists, today, self-driving cars are becoming reality with astonishing speed. We all hope that robots will make our lives better. After all, they have a nearly infinite capacity for processing information, their attention span is less limited than ours, and they make fewer mistakes. The scenario of self-driving cars, for example, offers new opportunities to facilitate the mobility of disadvantaged persons such as the disabled, the elderly, and persons too young to drive,

1

But what if a ride in a self-driving car ends in a fatal accident?[1] It would be foolish to assume that self-driving cars are error-proof. They can function properly only if they interact safely with their environment, especially surrounding traffic, which includes human drivers, pedestrians, cyclists and, in the future, other self-driving cars as well as "intelligent streets". A self-driving car may be unable to react appropriately to an unforeseen crisis, or its complicated technology may simply fail and cause an accident involving damage, injuries, or even the loss of human life. When this happens, there is bound to be a public outcry coupled with the demand for a determination of responsibility, including allocation of civil damages and criminal liability.[2]

The question of responsibility for such failures is difficult to resolve because self-driving cars not only operate without a driver, but also without an entirely defined operational modus. The defining feature of Intelligent Agents such as self-driving cars is smart technology, which mines data for patterns and reacts to them without human interference. Self-driving cars, for example, will most probably be capable of learning from the data they collect.[3] They detect patterns in the universe of data, build models from these patterns, and use these models to make predictions and/or take decisions.[4] This means that a self-driving car is in part the product of an engineer and in part a self-taught machine. It is a mash-up of pre-set algorithms, environmental conditions, the input of users, and artificial intelligence. Robots do, of course, follow instructions, but these instructions tell them to be independent, to learn from "experience", to try out new strategies, and to learn from the outcome of these trials. Since the machine draws its own conclusions, these outcomes cannot be predicted in advance. Neither its human producers and programmers nor its human users can foresee all of

---

[1] See Frank Douma & Sarah Aue Palodichuk, *Criminal Liability Issues Created by Autonomous Vehicles*, 52 Santa Clara L. Rev. 1157 – 1169 (1159) (2012); Eric Hilgendorf, *Können Roboter schuldhaft handeln?, Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, in Jenseits von Mensch und Maschine, 119 – 133 (Susanne Beck ed., 2012); Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 Md. L. Rev. 785 – 829 (791) (2015).

[2] For a discussion of criminalization in the context of new technologies, see Sarah Summers, *EU Criminal Law and the Regulation of Information and Communication Technology*, 3 Bergen L.J., 48 – 60 (54) (2015).

[3] For more information on technological approaches, see Adeel Lari, Frank Douma & Ify Onyiah, *Self-Driving Vehicles and Policy Implications: Current Status of Autonomous Vehicle Development and Minnesota Policy Implications*, 16 Minn. J.L. Sci. & Tech. 735, 745 (2015).

[4] Susanne Beck, *Dealing with Diffusion of Legal Responsibility: The Case of* Robotics, in Rethinking Responsibility in Science and Technology 167 – 181 (Fiorella Battaglia, Nikil Mukerji & Julian Nida-Rümelin eds., 2014).

a robot's possible actions; all these actors can say for certain is that the Intelligent Agent will devise its own solutions.[5]

Specialists in the areas of private law and insurance law have discussed the issue of liability for damages caused by the robot "itself".[6] Tort lawyers have suggested that Intelligent Agents should themselves be held liable for damages, possibly based on a no-fault liability scheme.[7] Other options would be to establish a private fund for indemnifying those who have suffered damage from robot action or to create insurance systems.[8] In private law, the possibility of holding entities other than natural persons liable for damages is nothing extraordinary; thus, the system can accommodate robot liability with no great difficulty.

For criminal law, by contrast, it is difficult to accommodate "acts" of non-human agents. Since there is no international criminal law governing robots, each national legal system will need to devise solutions based on its own general rules and principles. Pragmatic legal systems such as those of the United States may be able to accommodate their models of criminal responsibility to acts of robots much as they have integrated the "acts" of non-human, legal persons into their systems of criminal law.[9] But other legal systems are more hesitant in this respect. One example is Germany, whose criminal law doctrine is still firmly rooted in 19th century idealist philosophy. Germany has so far resisted the idea of expanding criminal liability to non-human agents, including corporations and other legal persons. So, if criminal responsibility for robot action can be integrated into the German system, it should be possible to replicate that feat anywhere. In the following we will therefore concentrate on the question of whether and how criminal liability for robots might fit into the German system of criminal law, and will contrast that with the attitude of U.S. law toward robot liability.

---

[5] For more scenarios of robots allegedly committing crimes, see Hartzog (note 1, supra), at 791.

[6] See Jochen Hanisch, *Haftung für Automation* (2010); Douma & Palodichuk (note 1, supra), at 1159; Elbert De Jong, *Regulating Uncertain Risks in an Innovative Society: A Liability* Law *Perspective,* in Robotik und Gesetzgebung 163–182 (Eric Hilgendorf & Jan-Philipp Günther eds., 2013).

[7] Cf. Hanisch (note 6, supra);Douma & Palodichuk (note 1, supra) at 1159; de Jong (note 6, supra); Peter Bräutigam & Thomas Klindt, *Industrie 4.0, das Internet der Dinge und das Recht,* Neue Juristische Wochenschrift 68, 1137 − 1142 (1138 − 1139) (2015); Lennart S. Lutz*, Autonome Fahrzeuge als rechtliche Herausforderung*, Neue Juristische Wochenschrift 68, 119 − 124 (2015).

[8] Malte-Christian Gruber, *Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs,* in Jenseits von Mensch und Maschine, 133 − 160 (Susanne Beck ed., 2012).

[9] E.g., Edmund P. Evans, *The Criminal Prosecution and Capital Punishment of Animals* <http://www.gutenberg.org/files/43286/43286-h/43286-h.htm> (last visited Nov. 16, 2015).

II. Criminal Liability of Robots?

Let us first look at the question of whether the law can treat Intelligent Agents as entities subject to criminal punishment.[10] In the past, this was primarily an issue for science fiction novels and movies.[11] But with the advent of autonomous, self-guided robots, it may well be that Intelligent Agents are no less "responsible" for their actions than human beings are.[12]

The issue of a robot's potential responsibility leads us back to the fundamental question of what it means to be a "person". Philosophers have long debated this question and have come to different conclusions. One approach has based personhood on the capacity for self-reflection. *John Locke*, for example, wrote that an "intelligent Agent", meaning a human person, must be "capable of a Law, and Happiness and Misery". Such an agent, moreover, must have a consciousness of his own past: "This personality extends itself beyond present existence to what is past, only by consciousness, – whereby it becomes concerned and accountable; owns and imputes to itself past actions, just upon the same ground and for the same reason as it does the present."[13] *Immanuel Kant* similarly emphasized the importance of self-consciousness:[14] since man is conscious of himself, he understands his own freedom and knows that his own will is the cause of each of his acts. For that reason, a person knows that he could have refrained from committing each of his acts and hence regards himself as responsible for his acts. *Kant* regards the irrepressible voice of conscience *(Gewissen)* as proof of the truth of this assertion. As a consequence, other persons may also hold the actor responsible for his acts.

---

[10] On this question, see Susanne Beck (note 4, supra), at 172 et seq.; Bert-Jaap Koops, Mireille Hildebrandt & David-Olivier Jaquet-Chiffelle, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, 11 Minn. J. L. Sci. & Tech., 497 – 561, (522 et seq.) (2010); Mireille Hildebrandt, *Autonomic and Autonomous "Thinking": Preconditions for Criminal Accountability*, in Law, Human Agency and Autonomic Computing 141 – 160 (Mireille Hildebrandt & Antoinette Rouvroy eds., 2011).

[11] E.g., in Star Trek, The Ultimate Computer (1968) McCoy declares: *"Compassion. That's the one thing no machine ever had. Maybe it's the one thing that keeps men ahead of them"* and Captain Kirk argues with a robot: *"There were many men aboard those ships. They were murdered. Must you survive by murder?"* M-5: *"This unit cannot murder."* Captain Kirk: *"Why?"* M-5: *"Murder is contrary to the laws of man and God."* Captain Kirk: *"But you HAVE murdered. Scan the starship Excalibur, which you destroyed. Is there life aboard?"* M-5: *"No life."* Captain Kirk: *"Because you MURDERED it. What is the penalty for murder?"* M-5: *"Death."* Captain Kirk: *"And how will you pay for your acts of murder?"* M-5: *"This – unit – must die."*

[12] Peter M. Asaro, *A Body to Kick, But Still No Soul to Damn,* in Robot Ethics: The Ethical and Social Implications of Robotics, 169 – 185 (180-183) (Patrick Lin, Keith Abney & George A. Bekey eds., 2011).

[13] John Locke, *An Essay Concerning Human Understanding*, § XVII No. 26 331 (1690).

[14] Immanuel Kant, *Kritik der praktischen Vernunft*, in Immanuel Kant, Werke in zehn Bänden, vol. 6, 223 (Wilhelm Weischedel ed., 1975).

Obviously, 18th and 19th century philosophers were not thinking about contemporary robots. But it is clear that a modern-day Intelligent Agent does not meet the requirements of personhood in the idealistic sense: While it may be able to learn and to make decisions that are unforeseeable by humans, a robot nonetheless cannot be conscious of its freedom, cannot understand itself as an entity with a past and a future, and certainly cannot grasp the concept of having rights and obligations.[15] Even robots that are able to learn do not have a conscience and do not reflect upon whether their actions are good or bad.[16] Therefore we do not view robots as "free" agents – which implies that we cannot hold them "personally" responsible for any harm they may cause.

To be sure, these statements refer to Intelligent Agents as we know them in 2016. Looking back at the rapid changes that have taken place in computer sciences in recent decades, it is not unlikely, however, that Intelligent Agents of the future will acquire qualities and capabilities that make them even more like humans. Should they gain the capacity for self-reflection and something like a conscience, the issue of their personhood may have to be rethought.[17]

But if robots are not persons in any philosophical or legal sense, can they perhaps still be punished?

1. United States

In many legal systems, corporations (clearly not "persons" in the sense discussed above) are subject to criminal sanctions.[18] In the United States, for example, corporate criminal liability has been in use for more than a century.[19] On the federal level and in some states, the doctrine of *respondeat superior* (borrowed from tort law) forms the basis of corporate criminal

---

[15] See Lawrence B. Solum, *Legal Personhood for Artificial Intelligences,* 70 N.C. L. Rev. 1231 (1992); Andreas Matthias, *Automaten als Träger von Rechten* (2d ed., 2010).

[16] See Koops et al. (note 10, supra), at 522 et seq.

[17] For views on the legal status of robots, see Susanne Beck, *Über Sinn und Unsinn von Statusfragen – zu Vor- und Nachteilen der Einführung einer elektronischen Person*, in Robotik und Gesetzgebung 239 – 262 (Eric Hilgendorf & Jan-Philipp Günther eds., 2013); Mireille Hildebrandt, *From Galatea 2.2 to Watson – And Back?,* in Human Law and Computer Law: Comparative Perspectives, 23 – 46 (27 et seq.) (Mireille Hildebrandt & Jeanne Gaakeer eds., 2013).

[18] In Switzerland, for instance, corporations and other legal entities can be prosecuted and fined up to five million Swiss francs if a wrongful act has been committed by individuals within the scope of business and cannot be traced back to a particular individual because the enterprise failed to adopt governance measures necessary to prevent criminal activity.

19 United States federal law allows corporations to be fined, sentenced to probation for up to five years, ordered to make restitution, and/or required to notify the public and their victims about their criminal wrongdoing. United States Sentencing Guidelines, Chapter 8. For an overview, see Pamela H. Bucy, *Corporate Criminal Responsibility,* in 1 Encyclopedia of Crime and Justice 259-265 (Joshua Dressler ed., 2d ed., 2002).

liability. This doctrine permits the imposition of criminal liability on corporations for acts committed by their agents on behalf of the corporation, for the benefit of the corporation, and within the scope of the agent's authority.[20] Several states of the United States have adopted a somewhat more restrictive approach based on the Model Penal Code, which – depending upon the crime involved – requires conduct of a "high managerial agent" for attribution to the corporation.[21] The Code, it should be noted, provides a "due diligence" defense for non-strict liability offenses. This defense applies if the corporate defendant proves that the high managerial agent with supervisory responsibility over the subject matter employed due diligence to prevent commission of the offense.[22]

2. Germany

As of yet, Germany has not recognized "genuine" criminal responsibility of legal entities.[23] The main objections raised by German authors might equally apply to the criminal liability of robots: first, legal entities cannot "act", i.e., they are not capable of autonomously setting goals for themselves and acting in order to reach those goals; and second, they are incapable of realizing the wrongfulness of their conduct and therefore cannot be blamed for their actions.

We shall leave open the question of whether these objections are apposite in respect of corporations; but do they apply to robots? The sight of a driverless car traveling on the road or of a drone gliding through the sky certainly gives the impression of Intelligent Agents as

---

[20] See, e.g., *New York Central & Hudson River Railroad v. United States,* 212 U.S. 481 (1909); *United States v. Singh*, 518 F.3d 236 (4th Cir. 2008); J. Kelly Strader, *Understanding White Collar Crime,* 16–17 (3d ed. 2011).

[21] MPC § 2.07. See Ellen S. Podgor, Peter J. Henning, Jerold H. Israel & Nancy J. King, *White Collar Crime*, 29 (2013).

[22] MPC § 2.07(5).

[23] An administrative fine can be imposed on corporations if one of their responsible agents has committed a criminal or administrative offense on the corporation's behalf (§ 30 Gesetz über Ordnungswidrigkeiten). For the German discussion on corporate criminal responsibility, see Klaus Volk, *Zur Bestrafung von* Unternehmen, JZ 48, 429 – 435 (1993); Günter Heine, *Die strafrechtliche Verantwortlichkeit von Unternehmen, Von individuellem Fehlverhalten zu kollektiven Fehlentwicklungen, insbesondere bei Großrisiken*, 1995; Günther Jakobs, *Strafbarkeit juristischer Personen?,* in Festschrift für Klaus Lüderssen zum 70. Geburtstag am 2. Mai 2002, 559 – 576 (Cornelius Prittwitz ed., 2002); Michael Hettinger (ed.), *Reform des Sanktionenrechts: Verbandsstrafe,* vol. 3 (2003); Bernd Schünemann*, Strafrechtliche Sanktionen gegen Wirtschaftsunternehmen?*, in Strafrecht und Wirtschaftsstrafrecht: Dogmatik, Rechtsvergleich, Rechtstatsachen, Festschrift für Klaus Tiedemann zum 70. Geburtstag, 429 – 448 (Ulrich Sieber, Gerhard Dannecker, Urs Kindhäuser, Joachim Vogel & Tonio Walter eds., 2008); Wolfgang Frisch, *Strafbarkeit juristischer Personen und Zurechnung,* in Gesamte Strafrechtswissenschaft in internationaler Dimension, Festschrift für Jürgen Wolter zum 70. Geburtstag am 7. September 2013, 349 – 374 (Mark A. Zöller, Hans Hilger, Wilfried Küper & Claus Roxin eds., 2013); Martin Böse, *Strafbarkeit juristischer Personen – Selbstverständlichkeit oder Paradigmenwechsel im Strafrecht*, ZStW 126, 132 – 167 (2014). See also the 2013 Draft Code of corporate criminal liability: Gesetzentwurf des Landes Nordrhein-Westfalen für ein Verbandsstrafgesetzbuch; ‹http://www.justiz.nrw.de/JM/justizpolitik/jumiko/beschluesse/2013/herbstkonferenz13/zw3/TOP_II_5_Gesetzentwurf.pdf›.

"actors", as masters of their own motion. But for German criminal law theorists, an "act" requires an autonomous will.[24] Whereas some writers speak of an "act" whenever the actor's movement is not dominated by someone else,[25] others regard the goal-orientation *(Finalität)* of a human act as its distinguishing characteristic. According to this latter view, a physical act can be subject to the criminal law only if it has been caused by a person's will to achieve a certain purpose.[26] Can robots fulfill this requirement? Due to their ability to adapt to their environment, Intelligent Agents can autonomously determine the steps they need to take in order to reach the goal that has been set for them. For example, a self-driving car may well be able to identify another vehicle approaching a junction and "decide" not to stop because it assumes from what it has learned that it has the right of way. Yet it is an open question whether an Intelligent Agent makes autonomous judgments that can be regarded as the product of its "own" will or whether it simply makes choices in accordance with its programming. Regardless of how this question is answered today, it cannot be ruled out that, in the future, Intelligent Agents, which even now are capable of learning, remembering, and selecting, will be ascribed the capability of forming their own will.[27]

But even if a robot is assumed to make judgments, it is – at least as of today – unable to set its own goals; a robot, moreover, is unaware of the *social* (let alone moral) consequences of its actions. There has been some debate as to whether an Intelligent Agent may someday acquire the ability to reflect on its own actions, but so far there has not been any positive evidence to that effect.[28] This means that the answer to the question of whether a robot can act ultimately depends on whether we employ a "thick" definition of an act, one that includes the ability to make *autonomous* decisions on goal-setting, or a "thin" definition, according to which an act is any physical movement that has not conclusively been determined by another agent.

---

[24] For an overview of various "act theories", see Hans-Heinrich Jescheck & Thomas Weigend, *Lehrbuch des Strafrechts, Allgemeiner Teil,* 217 – 231 (5th ed. 1996); Claus Roxin, *Strafrecht Allgemeiner Teil I: Grundlagen. Der Aufbau der Verbrechenslehre*, 236 – 271 (4th ed. 2006); Ingeborg Puppe, *vor § 13 marginal note 31 et seq.,* in Nomos Kommentar Strafgesetzbuch, (Urs Kindhäuser, Ulfrid Neumann & Hans-Ullrich Paeffgen eds., 4th ed. 2013).

[25] See Jürgen Baumann, Ulrich Weber & Wolfgang Mitsch, *Strafrecht Allgemeiner Teil*, 207 – 211 (11th ed. 2003).

[26] Hans Welzel, *Die deutsche strafrechtliche Dogmatik der letzten 100 Jahre und die finale Handlungslehre*, Juristische Schulung, 421 – 428 (1966); Hans Welzel, *Das deutsche Strafrecht*, 33 – 42 (11th ed. 1969); for a similar modern view, see Günter Stratenwerth & Lothar Kuhlen, *Strafrecht Allgemeiner Teil: Die Straftat*, 57 – 58 (6th ed. 2011); Günter Stratenwerth, *Schweizerisches Strafrecht. Allgemeiner Teil I: Die Straftat*, 128 – 129 (4th ed. 2011).

[27] Cf. Beck (note 17, supra); Hildebrandt (note 17, supra), at 27 – 28.

[28] See Douglas Hofstadter, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, 179 – 204 (1997).

Even if it is assumed that robots can act, the more difficult question remains of whether they can be blamed for what they do. In Germany, the "principle of blameworthiness" is deemed to be guaranteed by the constitution.[29] In a leading judgment of 1952, the Federal Court of Justice held that every adult person is capable of free and responsible moral self-determination. Since adult persons can decide in favor of the right and against the wrong, and can adapt their conduct to the requirements of the law, the Court continued, adult persons may be blamed for violating legal norms.[30] According to this reasoning, blameworthiness presupposes the actor's ability to decide between doing right and doing wrong or, in other words, presupposes his ability to avoid committing a wrongful act. This definition seems categorically to exclude the blameworthiness of even highly intelligent robots. In spite of their "intelligence", robots are machines that complete their tasks in accordance with their programming; they are not imbued with the ability of moral self-determination. This result does not change significantly even if we take into account certain results of experimental brain research that cast doubt upon the freedom of humans to decide between various courses of action on the basis of moral considerations.[31] Whatever neuroscientists may discover about the predetermination of human decision-making, social relations – including the formation and application of legal norms – are based on the assumption that man has a "free will", that is, that his actions can be attributed to him unless he is immature or mentally ill.[32]

---

[29] See *Entscheidungen des Bundesverfassungsgerichts* [hereinafter *BVerfGE*] 45, 187, 227; 95, 96, 140. In BVerfGE 123, 267, 413, the constitutional court maintained that even a constitutional amendment would not be able to abolish the principle of blameworthiness, because it is rooted in the supreme value of human dignity. On the principle of blameworthiness, see generally Roxin (note 23, supra), at 851 – 880; Stratenwerth & Kuhlen (note 25, supra), at 159 – 191; Kristian Kühl, *Strafrecht Allgemeiner Teil*, 364 – 430 (7th ed. 2012); Helmut Frister, *Strafrecht Allgemeiner Teil*, 29 – 40 (7th ed. 2015).

[30] *Entscheidungen des Bundesgerichtshofes in Strafsachen* [hereinafter BGHSt] 2, 194, 200.

[31] See, for example Wolf Singer, *Verschaltungen legen uns fest: Wir sollten aufhören, von Freiheit zu sprechen*, in Hirnforschung und Willensfreiheit: Zur Deutung der neuesten Experimente, 30 – 65 (Christian Geyer ed., 2004); Michael Pauen & Gerhard Roth*, Freiheit, Schuld und Verantwortung: Grundzüge einer naturalistischen Theorie der Willensfreiheit* (2008). For critical assessments from the perspective of criminal law, see Thomas Hillenkamp, *Strafrecht ohne Willensfreiheit? Eine Antwort auf die Hirnforschung,* Juristenzeitung 60, 313 – 320 (2005); Franz Streng, *Schuldbegriff und Hirnforschung*, in Festschrift für Günther Jakobs zum 70. Geburtstag am 26. Juli 2007, 675 – 692 (Michael Pawlik & Rainer Zaczyk eds., 2007); Gunnar Duttge, *Über die Brücke der Willensfreiheit zur Schuld – Eine thematische Einführung*, in Das Ich und sein Gehirn, Die Herausforderung der neurobiologischen Forschung für das (Straf-) Recht, 13 – 62 (Gunnar Duttge ed., 2009); Bettina Weißer*, Ist das Konzept strafrechtlicher Schuld nach § 20 StGB durch die Erkenntnisse der Neurowissenschaft widerlegt?, Goltdammers Archiv für Strafrecht 160, 26 – 38 (2013);* Reinhard Merkel, *Willensfreiheit und rechtliche Schuld: Eine strafrechtsphilosophische Untersuchung* (2d ed. 2014)*.

[32] See Günther Jakobs, *Strafrecht Allgemeiner Teil: Die Grundlagen und die Zurechnungslehre*, 484 – 485 (2d ed. 1991); Roxin (note 23, supra), at 869 – 880; Frister (note 28, supra), at 31 – 33. See also Immanuel Kant, *Kritik der reinen Vernunft*, in Immanuel Kant, Werke in sechs Bänden, vol. 2, 501 (Wilhelm Weischedel ed., 1956).

In the view of some German authors, criminal culpability has acquired a more functional character. They view attribution of culpability not as a moral reaction to a faulty exercise of free will but rather as a means to re-instate the community's trust in the validity of the law violated by the offender.[33] According to this view, the attribution of criminal culpability to a person presupposes that the person had the capacity to put into question the validity of a legal norm. Young children and insane persons do not have this capacity because their actions are not determined by reason and therefore cannot be seen as a challenge to the continuing validity of a legal norm. An actor's performance of a criminal act can be a threat to the continuing validity of a norm only if the actor in that particular situation can be regarded as participating in a discourse on normative questions. This, again, requires an actor who has the capacity to engage in self-reflection.[34] It would make little (social) sense to attribute culpability to a being that is incapable of recognizing its own past and evaluating its past actions in accordance with a moral reference system. An entity that does not have a conscience cannot participate in a dialogue on ethical issues and cannot respond to reproach. Such an agent – for example, a child or a person afflicted with a mental illness that severely impairs his ability of moral reasoning – can be physically restrained if he poses a threat to himself or others, but it makes little sense to treat him as culpable.

3. Robot perspective

Modern Intelligent Agents can make decisions based on an evaluation of their options. They can be taught to react to "moral dilemmas", that is, to choose to forego the pursuit of a goal if the goal can only be achieved by causing significant collateral harm. For example, while a self-driving car is programmed to reach its destination quickly and without detours, it can (and must) be programmed in such a way that it does not run over pedestrians blocking the road.[35]

---

[33] See Jakobs (note 32, supra), at 471 – 475.

[34] See Locke (note 13, supra); see further Kurt Seelmann, *Personalität und Zurechnung von der Aufklärung bis zur Philosophie des Idealismus*, in "Toujours agité – jamais abattu": Festschrift für Hans Wiprächtiger, 575 – 586 (Marianne Heer, Stefan Heimgartner, Marcel Alexander Niggli & Marc Thommen eds., 2011); Kurt Seelmann & Daniela Demko, *Rechtsphilosophie*, 152 – 154 (6th ed. 2014).

[35] The computer science community is currently discussing ethical dilemmas for Intelligent Agents: Should they be programmed to run over one pedestrian if by so doing they would save the lives of three others? See ‹http://marginalrevolution.com/marginalrevolution/2012/06/the-google-trolley-problem.html› (last visited Nov. 16, 2015). See also Torsten Pfützenreuter, *Intelligentes Missionsmanagement für autonome mobile Systeme* (2005).

Researchers engaged in the emerging field of machine ethics seek to determine whether – and if so, how – a system of decision-making based on morals can be programmed into robots.[36] At present, these efforts are still in their infancy, mostly because the formalization of ethical decisions appears to be exceedingly complicated.[37] Nevertheless, some researchers expect that Intelligent Agents will one day acquire the ability to engage in moral reasoning. Robots might be programmed with a system of "merits" and "demerits" for certain decisions they make, and that system could be treated as an analogue to human self-determination on moral grounds. Once that step has been taken, the attribution of criminal culpability to robots will no longer be out of the question.[38] The concept of culpability would then have to be adapted to the characteristics of Intelligent Agents. An Intelligent Agent could be "blamed" for its actions if it was able to recognize that they were undesirable and was hence able to evaluate them as "negative".

Assuming that robots can be blamed – can they also be punished? The sanctions in our criminal codes are geared toward human beings; they are neither meant nor designed to punish non-human entities.[39] Even where legal persons are formally subject to criminal punishment, the (mostly monetary) sanctions available are not intended to affect the corporation as such but rather to affect those human beings who have an interest in the financial well-being of the corporation: shareholders, members, or managers of the corporation.[40] It is difficult, at present, to imagine sanctions against Intelligent Agents that

---

[36] Thomas Georges, *Digital Soul: Intelligent Machines and Human Values* (2003); J. Storrs Hall, *Beyond AI: Creating the Conscience of the Machine* (2007); Wendell Wallach & Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (2009); *Machine Ethics* (Michael Anderson & Susan Leigh Anderson eds., 2011).

[37] See Nick Bostrom & Eliezer Yudkowsky, *The ethics of artificial intelligence*, in The Cambridge Handbook of Artificial Intelligence, 316 − 334 (Keith Frankish & William M. Ramsey eds., 2011); Luke Muehlhauser & Louie Helm, *Intelligence Explosion and Machine Ethics*, in Singularity Hypotheses: A Scientific and Philosophical Assessment, 101 − 126 (Amnon Eden, Johnny Søraker, James H. Moor & Eric Steinhart eds., 2012) ‹http://intelligence.org/files/IE-ME.pdf› (last visited Nov.16, 2015); Christopher Charles Santos-Lang, *Moral Ecology Approaches to Machine Ethics,* in Machine Medical Ethics, 111 − 127 (Simon Peter van Rysewyk & Matthijs Pontier eds., 2014).

[38] Cf. Malte-Christian Gruber, *Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs*, in Jenseits von Mensch und Maschine, Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 133 − 160 (150) (Susanne Beck ed., 2012); Gunther Teubner, *Elektronische Agenten und grosse Menschenaffen: Zur Ausweitung des Akteurstatus in Recht und Politik,* Zeitschrift für Rechtssoziologie 27, 5 − 30 (2006); Susanne Beck & Benno Zabel, *Person, Persönlichkeit, Autonomie − Juristische Perspektiven,* in Persönlichkeit: Neurowissenschaftliche und neurophilosophische Fragestellungen, 49 − 82 (Michael Zichy & Orsolya Friedrich eds., 2014); Mireille Hildebrandt, *Criminal Liability and ´Smart´ Environments,* in Philosophical Foundations of Criminal Law, 507 − 532 (R.A. Duff & Stuart P. Green eds., 2014).

[39] See Wallach & Allen (note 36, supra); Matthias (note 15, supra).

[40] It is difficult to answer the question whether and under what conditions it is appropriate to punish these persons for the wrongdoing of others whose acts have been attributed to the legal person; see Mordechai Kremnitzer & Khalid Ghanayim, *Die Strafbarkeit von Unternehmen,* Zeitschrift für die gesamte

would fulfil the same purposes as criminal sanctions imposed on human beings. Since a robot does not have property (at least, it cannot be aware that it has property),[41] any fine imposed on it would have to be paid by its legal owner or some fund created for potential liability. The physical destruction or impairment of the robot might, to a human observer, look like corporal punishment or even the death penalty; but it would not have a comparable effect on the robot, at least as long as it is not imbued with a will to live (in "good health"). Robots, in short, are incapable of understanding the meaning of punishment and can therefore not draw a connection between anything "done to them" and their prior fault.

In sum, robots that take action only in accordance with the way they have been programmed may not qualify as actors in the sense of the criminal law. Even robots that appear to make their own decisions by identifying patterns in their memory and selecting one possible course of action currently do not possess the sort of self-awareness and self-reflection that would make them possible targets of blame. But this might change if Intelligent Agents were to learn to address and decide issues of morality.[42] The final step would be to make robots capable of understanding punishment, that is, to teach them to associate certain changes in their environment with the wrongfulness of their prior acts. But it seems that there is a long way to go before this final stage is reached.

III. Criminal Liability of the "Human behind the Machine"?

If robots cannot be punished, under what conditions should humans be held criminally responsible for producing, programming, or using intelligent machines that cause harm? For example, should the engineer who helps develop a self-driving car incur criminal liability for homicide if the car "makes a decision" that causes the death of a bystander?

Strafrechtswissenschaft 113, 539 – 564 (542 et seq.) (2001); Kurt Seelmann, *Unternehmensstrafbarkeit: Ursachen, Paradoxien und Folgen*, in Wirtschaft und Strafrecht, Festschrift für Niklaus Schmid zum 65. Geburtstag, 169 – 186 (Jürg-Beat Ackermann, Andreas Donatsch & Jörg Rehberg eds., 2001); Matthias Forster, *Die strafrechtliche Verantwortlichkeit des Unternehmens nach Art. 102 StGB*, 38 – 39 (2006); Thomas Weigend, *Societas delinquere non potest?: A German Perspective*, Journal of International Criminal Justice 6, 927 – 945 (941-942) (2008); Gerson Trüg, *Sozialkontrolle durch Strafrecht – Unternehmensstrafrecht*, StrafverteidigerForum, 471 – 485 (473) (2011).

[41] See Koops et al. (note 10, supra), at 527 et seq.

[42] Mireille Hildebrandt, *From Galatea 2.2 to Watson – And Back?*, in Human Law and Computer Law: Comparative Perspectives, 23 – 46 (27 et seq.) (Mireille Hildebrandt & Jeanne Gaakeer eds., 2013); Susanne Beck, *Dealing with Diffusion of Legal Responsibility: The Case of* Robotics, in Rethinking Responsibility in Science and Technology, 167 – 181 (Fiorella Battaglia, Nikil Mukerji & Julian Nida-Rümelin eds., 2014).

If a human intentionally or knowingly programs a robot so that it causes harm to a person, the programmer's criminal responsibility can easily be established on the basis of traditional concepts of attribution and *mens rea*: The programmer commits the criminal act by using the robot – irrespective of its artificial intelligence – as a tool for carrying out the programmer's intention, and she does so with the requisite intent or knowledge.[43]

Cases of negligence pose greater problems. Can the producer, programmer, and/or owner – collectively referred to in the following as "operator" – of a self-driving car be held criminally responsible for negligent bodily injury to the pedestrian? Take, for example, a self-driving car that runs over a small child because its environment-scanning sensors misinterpreted its surroundings and failed to identify the child as a human being.[44] Or, take the unfortunate experience of a South Korean woman who learned the hard way about the risks of employing a robot: while she lay napping on the floor, her robot vacuum cleaner sucked up her hair, apparently mistaking her mane for dust.[45]

To answer this question and comparable questions of liability arising from a whole host of scenarios, we can refer to a body of law – statutes and cases that deal with the (negligent) use of machines by human beings.[46] We need not, in principle, depart from the conventional tools of products liability. In the South Korean case, it would appear that one or more of the robot vacuum cleaner's pre-set algorithms – possibly the one controlling its ability to sense and avoid obstacles – contributed to its failure to recognize the sleeping woman's hair as an obstacle, a failure that led to an injury.

1. Germany
Under German law, an actor is liable for criminal negligence if causation of a relevant harm (e.g., someone's death or injury) can be attributed to the actor, if he could have foreseen the

---

[43] Another case that falls into this category is the so-called Google bomb, which manipulates search results for individual websites.

[44] For a comprehensive overview on differences in current sensor technologies, see Adeel Lari, Frank Douma & Ify Onyiah, *Self-Driving Vehicles and Policy Implications : Current Status of Autonomous Vehicle Development and Minnesota Policy Implications*, 16 Minn. J.L. Sci. & Tech. 735 – 769 (745) (2015).

[45] Justin McCurry, *South Korean Woman's Hair "Eaten" by Robot Vacuum Cleaner as She Slept,* The Guardian*, Feb. 9, 2015,* http://www.theguardian.com/world/2015/feb/09/south-korean-womans-hair-eaten-by-robot-vacuum-cleaner-as-she-slept (last visited Nov. 16, 2015).

[46] Bräutigam & Klindt (note 7, supra), at 1140; Bryant Walker Smith, *Proximity-Driven Liability*, 102 Geo. L.J., 1777 – 1820 (2014); Stephen S. Wu, *Product Liability Issues in the U.S. and Associated Risk Management*, in Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte, 575 – 592 (Markus Maurer, J. Christian Gerdes, Barbara Lenz & Hermann Winner eds., 2015).

harm, and if he failed to exercise the due care necessary to avert the foreseeable harm.[47] The general idea behind responsibility for negligence is the actor's failure to pay sufficient attention to the harm he may cause to others. Liability is imposed if the actor could have prevented the harm in question if he had applied himself to doing so. Where even a diligent person could not realize that harm is imminent, there is no wrongdoing and hence no punishability. For example, if a wire is invisibly connected to an explosive device and anyone who touches the wire triggers the device, a person who causes the explosion by touching the wire cannot be punished as long as she had no reason to suspect that her seemingly harmless act would cause an explosion. Similarly, if a person is unable to prevent the occurrence of harm, for example, a train driver who is unable to stop the train in time to avoid hitting a child playing on the tracks, that person cannot be held criminally liable for negligence.

The standards of due attention and due care are geared toward human beings. They cannot simply be transferred to robots because robots cannot "foresee" consequences they have not been programmed to foresee. But what about the human being who designs, programs, or produces the robot?

Intelligent Agents of the latest generation, as we have seen, make decisions that their operators cannot foresee in detail. A self-driving car, for example, must interact with an unpredictable and only partly smart environment. Depending on its built-in learning-mode, the robot will use all stored and incoming information to establish working predictions regarding its dynamic environment. Operators know that robots will independently analyze the information they acquire and that they will act autonomously in response to the results of their analyses.

This means that the operator cannot reduce to zero the possibility that robots may cause harm to others. This fact suggests two mutually-exclusive conclusions as to the operator's liability for negligence. It could be argued that he cannot be held responsible because the machine is acting "on its own"; alternatively, it could be claimed that he can foresee any and all harm that robots might cause and therefore should face *de facto* strict liability for the results of the robots' acts.

---

[47] Gunnar Duttge, *§ 15 Rdn. 107 et seq.*, in Münchener Kommentar zum Strafgesetzbuch (Wolfgang Joecks & Klaus Miebach eds., 2d ed. 2011). For considerations on liability for negligence concerning Intelligent Agents, see Eric Hilgendorf, *Grundfragen strafrechtlicher Compliance am Beispiel der strafrechtlichen Produkthaftung für teilautonome technische Systeme,* in Criminal Compliance vor den Aufgaben der Zukunft, 19 – 32 (24 et seq.) (Thomas Rotsch ed., 2013).

The first line of argument is unconvincing. The fact that Intelligent Agents are generally unpredictable cannot relieve their operators of liability because it is the robots' very unpredictability that gives rise to duties of care. Likewise, if the manager of a zoo releases a tiger from its cage and the tiger kills people on the street, the zoo manager could not successfully argue that tigers are wild animals and therefore cannot be controlled. Since we have seen that robots cannot be held criminally liable, generally exempting their operators from liability would mean, in effect, that no one would be held criminally responsible for the death of a random victim of an errant driverless car.[48] Therefore, a person who can foresee that his action might harm interests protected by criminal law (such as the life and health of other persons) is obliged to refrain from that action.[49] Hence, if the zoo manager can foresee that the tiger, if set free, will harm human beings, he must refrain from releasing the tiger from its cage. The same applies to potentially dangerous products: If the producer of a car can, with appropriate diligence, know that the vehicle's brakes are unreliable in bad weather, the producer violates her duty of care if she nevertheless markets the car.

Germany has developed stringent rules of civil liability for unsafe products, and these rules have largely been transferred into criminal law, so that their breach may lead to criminal responsibility for negligence.[50] Before marketing a product, the producer must ascertain that it meets current scientific and technical standards and that its safety for customers has been adequately tested.[51] When the product is on the market, the producer must continuously monitor customer feedback and must react immediately to complaints of accidents or harm caused by the product. If necessary to avoid further damage, the producer must issue warnings, recall defective products for repair, or even stop marketing them altogether. If the producer fails to adhere to these standards, he may be criminally liable for any harm caused

---

[48] For the problems of civil liability under U.S. law, see e.g. John W. Terwilleger, *Navigating the Road Ahead: Florida's Autonomous Vehicle Statute and Its Effect on Liability,* 89 Fla. B.J. 26 – 37 (27) (2015).

[49] Duttge (note 47, supra), *§ 15 Rdn. 121 et seq.*

[50] The leading case on criminal product liability is BGHSt 37, 106 ("*Lederspray*"). On the relationship between criminal and civil product liability, see Lothar Kuhlen, *Grundfragen der strafrechtlichen* Produkthaftung, Juristenzeitung 49, 1142 – 1146 (1994); Winfried Hassemer, *Produktverantwortung im modernen Strafrecht*, 74 (2d ed. 1996); Joachim Vogel, *vor § 15 marginal note 277 – 281*, in Strafgesetzbuch. Leipziger Kommentar (Wilhelm Heinrich Laufhütte, Ruth Rissing-van Saan & Klaus Tiedemann eds., 12th ed. 2007); Sabine Gless, *Strafrechtliche Produkthaftung*, recht, 54 – 64 (57) (2013); Lothar Kuhlen, *Strafrechtliche Produkthaftung*, in Handbuch Wirtschaftsstrafrecht, 79 – 114 (Hans Achenbach & Andreas Ransiek eds., 4th ed. 2015).

[51] See for IT products: Bräutigam & Klindt (note 7, supra, at 1140).

by the product and may be convicted of intentional (if he was aware of the risk) or negligent bodily injury or homicide by omission.[52]

The reason for this type of criminal product liability is not the *unlawful* creation of a risk but the mere fact that the producer, in pursuing economic interests, *lawfully* creates a risk for the general public by releasing an Intelligent Agent whose reactions cannot be safely predicted and controlled.[53] The unique feature of this concept of criminal liability is the fact that a perfectly legal act – the marketing of a self-driving car in accordance with the current state of knowledge and technology – may trigger criminal liability for omission.[54] It may be difficult for the car industry to accept this broad ambit of liability. But victims of accidents caused by malfunctioning self-driving cars would find it equally difficult to accept a situation in which, in the absence of a driver, no one is held responsible for the damage caused.

2. United States

The situation is similar in the United States. If the victim of the robot vaccum cleaner mentioned above brought suit against the robot's operator in the United States, in order to recover she would have to prove that the robot's failure was the result of a manufacturing defect, a design defect, of inadequate instructions or warnings regarding the risks of harm posed by the robot, or of some combination of these three categories.[55] Furthermore, if the facts indicated that the manufacturer took a substantial and unjustifiable risk of causing the injury, the manufacturing enterprise could also face liability for criminal negligence, defined as a gross deviation from the standard of reasonable care.[56] Where liability is formally based on fault, manufacturers must take all reasonable steps proportionate to the risk to reduce foreseeable risks of harm. In most instances, this means that manufacturers must do the following: exercise reasonable care to eliminate all substantial dangers from their products

---

[52] Thomas Weigend, *§ 13 Rdn. 53 et seq.*, in Strafgesetzbuch. Leipziger Kommentar (Wilhelm Heinrich Laufhütte, Ruth Rissing-van Saan & Klaus Tiedemann eds., 12th ed. 2007)*;* Kuhlen (note 50, supra), at 79 – 114.

[53] See for Switzerland: Amtliche Sammlung der Entscheidungen des Schweizerischen Bundesgerichts [BGE] 134 IV 255 (260 et seq.); for Germany: Lothar Kuhlen, *Strafhaftung bei unterlassenem Rückruf gesundheitsgefährdender Produkte – Zugleich Anmerkung zum Urteil des BGH vom 6.7.1990 – 2 StR 549/89*, Neue Zeitschrift für Strafrecht 10, 566 – 570 (568) (1990)*;* Jakobs (note 32, supra), at 812 – 813; Georg Freund*, Erfolgsdelikt und Unterlassen*, 199 – 223 (1992).

[54] As to liability for omission based on unlawful creation of risk, see Günter Stratenwerth, *Schweizerisches Strafrecht. Allgemeiner Teil I: Die Straftat*, 467 – 468 (4th ed. 2011); Andreas Donatsch & Brigitte Tag, *Strafrecht I. Verbrechenslehre*, 308 (9th ed. 2013); Frister (note 29, supra), at 303 et seq.

[55] Restatement (Third) of Torts: Products Liability § 2 (1998); Douma & Palodichuk (note 1, supra), at 1159; Bruce L. Ottley, Rogelio A. Lasso, & Terrence F. Kiely, *Products Liability Law,* 123 – 124 (2d ed. 2013).

[56] Joshua Dressler, *Understanding Criminal Law,* 132 (7th ed. 2015). For a discussion of corporate criminal negligence, see David Kerem, *Change We Can Believe In: Comparative Perspectives on the Criminalization of Corporate Negligence*, 14 Transactions: Tenn. J. Bus. L., 95 – 115 (2012).

that can reasonably be designed away; warn consumers about all substantial hidden dangers that remain; produce their products in such a way as to minimize dangerous manufacturing flaws; and be careful to avoid misrepresenting the safety of their products.[57] In order to avoid negligence-based liability, manufacturers must exercise reasonable care in all of these respects; if they do so, responsibility for any remaining dangers in the use of the products – whether defects, inherent hazards, or generic risks – must be borne by consumers. In the context of strict liability, manufacturers of unavoidably hazardous products must ensure that their products are free of production defects and must warn consumers of hidden dangers.[58] While more than thirty states have adopted some version of duties arising after the sale of a product, including the duty to warn and the duty to update,[59] post-sale duties are not uniform and consistent among the states of the United States, and courts have repeatedly refused to impose on manufacturers a post-sale duty to update (a duty that has been rejected by scholars as well).[60]

IV. Limiting Criminal Liability for Negligence

There is no good reason why the same standards should not apply, in principle, to the marketing of robots as well. One should indeed beware of giving profit-seeking operators *carte blanche* for taking inappropriate risks to the life and health of other persons. Yet, there may be good reasons for limiting producers' liability with regard to self-driving cars. If the use of self-driving cars leads to an overall reduction of accidents and provides the elderly and disabled with equal access to the advantages of personal mobility, it would seem unfair to place the (criminal) responsibility for causing harm entirely with the operator of the vehicle. Moreover, if the risk of criminal liability could not be averted even by exercising the greatest care possible, robots would soon be eliminated from the market, and the benefits associated with their introduction would be lost. Whereas it is true that operators of self-driving cars create risks to the life and health of others, one should not forget that the same is true for the

---

[57] Merton E. Marks, *US Product Liability Law*, International Business Lawyer 02, 69 – 72 (1998).
[58] David G. Owen, *Inherent Product Hazards*, 93 Ky. L.J. 377 – 422 (379) (2004).
[59] Tom Stilwell, *Warning: You May Possess Continuing Duties After the Sale of Your Product!*, 26 Rev. Litig. 1035 – 1066 (1037) (2007).
[60] Bryant Walker Smith (note 46, supra) at 1802 – 1808.

manufacturing and sale of traditional, person-driven cars. In 2014, 3,377 persons died on German streets as a result of traffic accidents, most of which were caused by human failure.[61]

Replacing human-driven cars by self-driving cars might indeed reduce the overall harm caused in street traffic; society may therefore have a valid interest in promoting the use of self-driving cars. Rather than devising all-or-nothing rules, it is necessary to carefully balance the conflicting interests involved. There is, on the one hand, the interest of victims of robots' malfunctioning in receiving redress for the harm they have suffered. On the other hand, producers have a legitimate interest in profitably marketing robots whose development they have bankrolled. Society at large is interested in reliable and convenient technological applications but also in technological progress, including the benefits that self-driving cars can provide. Given these conflicting interests, we should look for an intermediate solution with regard to criminal responsibility for negligently causing harm. Although the unpredictability of the results of autonomous learning makes the possibility of harmful activity of robots foreseeable, we might wish to limit their operators' criminal liability by adjusting the standard of care and introducing a margin of tolerance of some errors in designing and programming such cars.

A reduction of criminal responsibility could be achieved by limiting attribution of harmful results. Criminal law theory typically does not consider an actor who causes a harmful result responsible if his conduct and the resulting harm are not linked in a way that sustains an attribution of responsibility. One reason for excluding responsibility in this situation could be the primary attribution of the harmful result to the autonomous act of another person, to the victim himself, or to "chance" (which is another word for normal risks inherent in living in our dangerous world).

Normally, harm caused by a person's careless act or omission is attributed to that person. But traditional German doctrine[62] as well as U.S. scholars[63] recognize an exception for cases in which another person autonomously interferes with a causal chain of events, affecting it in such a way that the harmful result no longer appears to be the "work" of the original actor.

---

[61] Statistisches Bundesamt, https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/TransportVerkehr/Verkehrsunfaelle/Tabellen/UnfaelleVerunglueckte.html (last visited Nov. 16, 2015).

[62] *See, e.g.,* Hans-Heinrich Jescheck & Thomas Weigend, *Lehrbuch des Strafrechts, Allgemeiner Teil,* 281 (5th ed. 1996); Puppe (note 24, supra), *vor § 13 marginal notes 236 – 238,*; Jörg Eisele, *vor § 13 StGB marginal notes 77, 100 et seq.*, in Strafgesetzbuch, Kommentar (Adolf Schönke & Horst Schröder eds., 29th ed. 2014).

[63] Wayne R. La Fave, *Criminal law,* § 6.4(f)(5), 367 (5th ed. 2010); Sanford H. Kadish, Stephen J. Schulhofer & Carol S. Steiker, *Criminal Law and Its Processes: Cases And Materials*, 586 – 607 (9th ed. 2012).

For example, if A culpably causes a traffic accent that injures victim V, and V is taken to the hospital where he dies of his injuries, V's death is attributed to A's act. But it may well be different if B sets a fire to the hospital building while V is being treated there, and V dies in the flames. In this case, V's death will not be attributed to A's act, even though V would not have died in the hospital unless A had injured him. The limits of attribution in cases where a third person intervenes are far from clear, however.[64] There is no absolute protection of a merely negligent first actor[65] from attribution of guilt, even if the injury was actually brought about by a third person.[66]

In our context, one would have to answer the question of whether the intervention of a robot breaks the chain of attribution between the operator's act and the harm caused. For an observer, the sudden swerving of a self-driving car onto the sidewalk may appear to be an arbitrary, autonomous decision of the car and not the result of faulty programming or some other negligent act of the person who produced or marketed the car. But there are two strong arguments against categorically relieving the "man behind the machine" of responsibility by cutting off the chain of attribution. First, appearances may be misleading – the harm caused may indeed be the result of negligent programming and not of some unforeseeable quirk of the car. Second, as long as the robot itself cannot be held criminally responsible, the victim (and society) may face a responsibility gap: it would be impossible to hold either the machine or the persons behind it criminally liable, even for serious harm caused by the robot. Such a responsibility vacuum might cause a precipitous drop in support for robotic inventions. These considerations counsel against generally absolving the robot's operators of responsibility for harm caused by the robot.

According to another aspect of the theory of attribution, a person is not criminally responsible if the harm at issue is the result of a "normal" risk of daily life; in such situations the victim is expected to bear the harm without redress. For example, if D persuades V to take a walk in

---

[64] José Antonio Caro John, *Das erlaubte Kausieren verbotener Taten – Regressverbot* (2007); Puppe (note 24, supra), *vor § 13 marginal note 167 – 168;* Joshua Dressler, *Understanding Criminal Law*, 190 ("There are no hard and fast rules for rendering the commonsense 'community justice' determination of when an intervening cause supersedes the defendant's conduct.") (7th ed. 2015).

[65] As in the case where A negligently leaves a loaded gun in a public place and B grabs the gun and shoots V. On the argument that an intentional act of B should always preclude A's liability for negligence (so-called *Regressverbot*), see Wolfgang Naucke, *Über das Regressverbot im Strafrecht*, ZStW 76, 409 – 440 (1964); Wolfgang Frisch, *Tatbestandsmäßiges Verhalten und Zurechnung des Erfolges*, 62 – 63 (1988); Roxin (note 24, supra), at 236 – 237; Caro John (note 64, supra)*,* at 55 – 56; Puppe (note 23, supra), *vor § 13 marginal note 167 – 168, 236 – 237*; Eisele (note 62, supra), *vor § 13 StGB marginal note 77, 100 et seq*.

[66] If A runs over V with his car, intending to kill him, V is only injured and dies in the hospital fire set by B, V's death may still not be attributed to A, but the case is much closer than if A was only negligent.

the forest, and a tree cracks from natural causes, killing V, the death of V will not be attributed to D even though V would not have been killed without D's persuasion to take a walk. This result holds true, according to German doctrine,[67] even if D had wished V to be killed during the walk. In the United States, a so-called coincidental intervening cause, such as a falling tree or a bolt of lightning, likewise breaks the chain of causation caused by a defendant's criminal act, unless the intervening cause is foreseeable.[68] Thus, if – as in the above example –the death of V caused by a falling tree cannot be attributed to D, even if D secretly wished for V's death, D commits no criminal act at all.

Relying on these theories, it might be argued that robots are part of the "normal" risks of life, comparable to lightning or falling trees. At present, however, and in the near future, the appearance of a potentially dangerous robot cannot be regarded as an everyday occurrence that members of the general public must simply tolerate and integrate into their lives. Rather, releasing a robot into the real world is still viewed as creating an *exceptional* risk, so that any harm caused by the robot would normally be attributed to its operator if causation can be proved.

But this may change in the foreseeable future. In the internet, the ubiquity of search engines engaged in data processing has arguably made them a commonplace occurrence, with the consequence that their inherent risks are accepted by most users. Indeed, users who order goods and services from online merchants and who look for information with the help of search engines are fully aware of the fact that the Intelligent Agents operating these internet services are storing their personal data and may later use these data in violation of users' personality rights.[69] It may therefore no longer be possible to attribute the ensuing harm to personal rights to the operator of the Intelligent Agent. It may well be that in the future self-driving cars will similarly become "normal" partners of interaction. As soon as they are viewed as an everyday part of street traffic, encountering such vehicles will become a normal risk for humans, and the operator of the self-driving car will not be criminally liable for its

---

[67] See Tonio Walter*, vor § 13 StGB, marginal notes 103 et seq.,* in Strafgesetzbuch. Leipziger Kommentar (Wilhelm Heinrich Laufhütte, Ruth Rissing-van Saan & Klaus Tiedemann eds., 12th ed. 2007); Puppe (note 24, supra), *vor § 13 marginal notes 236 et seq.*

[68] Kadish, Schulhofer & Steiker (note 63, supra), at 571 – 584; Dressler (note 47, supra), at 191 – 193.

[69] See Naren Ramakrishnan, Benjamin J. Keller & Batul J. Mirza, *Privacy risks in Recommender Systems*, available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.2932&rep=rep1&type=pdf> (last visited Nov. 16, 2015); for more detailed information, see Krista Nadakavukaren Schefer, *Ein völkerrechtlicher Schutz der kollektiven Privatsphäre? Der Schutz der Privatsphäre und die Anonymität im Zeitalter kommerzieller Daten,* Zeitschrift für Schweizerisches Recht, 259 – 288 (2014).

generally foreseeable malfunctioning but only for harm incurred due to preventable construction, programming, or operating errors.

Even before this time arrives, however, we might choose to limit criminal liability of operators by reducing their duty of care,[70] in particular with regard to the rule that one must not expose others to risks that one cannot control. Even where it is foreseeable that robots may commit fatal mistakes due to misguided self-learning, the duty of operators of (at least some) robots could be reduced to employing the best knowledge and technology available in manufacturing, programming, testing and monitoring them. Producers of robots must test them extensively and must closely observe, in particular, their autonomous learning processes. They must, moreover, monitor feedback from customers and must react immediately to reports of harmful conduct.[71] If, for example, a newly introduced self-driving car for unknown reasons has isolated incidents of malfunctioning, the producer will have to examine possible causes. If these incidents cannot be explained by improper handling or interference of third parties, and if the problem cannot be resolved by re-programming the car, the producer will have to take the car off the market. If the producer fails to do so and the car causes harm to persons, he will be liable for negligent or even intentional bodily injury or homicide.[72]

Operators who comply with these strict standards could be deemed to have fulfilled their duty of care, even though they (along with everyone else) know that certain risks remain. If a robot, due to its autonomous learning process, deviates from what was expected, the harm caused by such (generally foreseeable) aberrations would be attributed to "society", which agrees to accept certain residual risks necessarily associated with the employment of Intelligent Agents. The person harmed by such malfunctioning would consequently be regarded as a victim of a socially accepted risk, not as a victim of the negligent wrongdoing of any particular person.

---

[70] Cf. Vogel (note 50, supra) *vor § 15 marginal notes 214 et seq.;* Puppe (note 24, supra), *vor § 13 marginal notes 157 et seq.;* Eisele (note 62, supra), *Vorbem. §§ 13 ff. StGB, marginal note 91*. An alternative and functionally equivalent solution might be to recognize responsible risk-taking (*erlaubtes Risiko*) as a grounds of justification; see the discussion in Roxin (note 24, supra), at 382 – 383; Theodor Lenckner & Detlev Sternberg-Lieben, *Vorbem. §§ 32 ff. StGB, marginal note 107b*, in Strafgesetzbuch, Kommentar (Adolf Schönke & Horst Schröder eds., 29th ed. 2014). But it might be difficult to explain to the victim of an out-of-control self-driving car that the harm done to him was justified and that, therefore, the victim must rightfully suffer it without redress.

[71] Kuhlen (note 54, supra).

[72] Sabine Gless (note 52, supra), at 59, has called this a "dormant duty to act", which comes alive whenever reports of harmful incidents appear.

V. Conclusion: The Policy Issue in Germany and the United States

The prospect of totally autonomous cars being part of daily life may still be some years away. In Germany, technology capable of driving a car, such as auto-pilot or parking devices, may only be used if a human being present in the car is capable of taking over the steering function in case of an emergency. In the United States, where self-driving cars currently undergo testing,[73] policy guidelines and legislation require a properly licensed driver to be present in the vehicle and to be ready to intervene if a technical problem arises.[74] The potentially uneasy co-existence of humans and robots raises the question of the extent to which the human on board may rely on the self-driving car and under what circumstances he must take over. Practical questions such as these may be of greater urgency to today's lawyers than the futuristic ones treated in this article; but they can more easily be resolved within the confines of traditional criminal law doctrine.

In contrast, totally autonomous cars and other Intelligent Agents push traditional criminal law doctrine to its limits. We have seen that, at present, self-driving cars cannot be held criminally liable because they lack the ability to act culpably (in any traditional sense) and because they cannot comprehend the idea of punishment. In the future, however, robots may become so similar to human beings that they, like us, will be able to "feel" the effects of criminal punishment. Once this stage has been reached, it might well make sense to consider punishing robots.

Both in Germany and the United States, operators of self-driving cars may be convicted of negligent injury or homicide if a car causes injury or death due to a foreseeable malfunction. In that case, criminal liability can be imposed on operators in accordance with general rules, even if their negligence consists only in a failure to react properly to reports of dangerous incidents. But foreseeing and controlling the conduct of self-teaching robots may become so difficult for human operators that we may need to raise the requirements for criminal negligence liability, lest the risk of punishment inhibit the further development and use of robots. If society embraces the convenience, the opportunities, and the safety assurances associated with self-driving cars, it should also be willing to accept the fact that (possibly

---

[73] E.g., <http://www.google.com/selfdrivingcar> (last visited Nov. 16, 2015).

[74] E.g., Policy of U.S. Department of Transportation on Automated Vehicle Development (May 30, 2013), available at <http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf> (last visited Nov. 16, 2015); Nev. Rev. Stat. Ann. § 482A.070 (2014); Cal. Veh. Code § 38750 (b) (2015); Fla. Stat. 316.86(1).

very rare cases of) unexpected actions of robots will lead to (generally) foreseeable harm to random victims.

Tolerance for robot malfunctions must however be subject to strict limitations. The challenge remains to strike a fair balance between society's interest in promoting innovation and the dangers associated with the use of Intelligent Agents with destructive potential. One factor to be considered in the balancing process is the social benefit of the robot at issue in relation to its potential for harm.[75] Under this kind of utilitarian approach, standards of care should be stricter with respect to robots that are of lesser social value, such as drones or toys. With respect to self-driving cars, on the other hand, the risk remaining after careful testing and monitoring may be offset against the general benefits of using such cars.

In the end, each society must answer for itself the question of whether investment in chances of a better life should be rewarded with an exemption from criminal responsibility for some of the risks involved – and what these risks are.

---

[75] Cornelius Prittwitz, *Strafrecht und Risiko*: *Untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft*, 297 – 298 (1993); Christian von Bar, *Die Grenzen der Haftung des Produzenten*, in Produktverantwortung und Risikoakzeptanz, 40 – 41 (Manfred Lieb ed., 1998).