Killing by Autonomous Vehicles and the Legal Doctrine of Necessity

Santoni de Sio, Filippo

DOI
10.1007/s10677-017-9780-7

Publication date
2017

Document Version
Final published version

Published in
Ethical Theory and Moral Practice: an international forum

Citation (APA)
Santoni de Sio, F. (2017). Killing by Autonomous Vehicles and the Legal Doctrine of Necessity. *Ethical Theory and Moral Practice: an international forum*, *20*(2), 411-429. https://doi.org/10.1007/s10677-017-9780-7

Ethic Theory Moral Prac (2017) 20:411–429
DOI 10.1007/s10677-017-9780-7

Killing by Autonomous Vehicles and the Legal Doctrine of Necessity

Filippo Santoni de Sio [1]

Accepted: 17 January 2017 / Published online: 22 February 2017

Abstract How should autonomous vehicles (aka self-driving cars) be programmed to behave in the event of an unavoidable accident in which the only choice open is one between causing different damages or losses to different objects or persons? This paper addresses this ethical question starting from the normative principles elaborated in the law to regulate difficult choices in other emergency scenarios. In particular, the paper offers a rational reconstruction of some major principles and norms embedded in the Anglo-American jurisprudence and case law on the "doctrine of necessity"; and assesses which, if any, of these principles and norms can be utilized to find reasonable guidelines for solving the ethical issue of the regulation of the programming of autonomous vehicles in emergency situations. The paper covers the following topics: the distinction between "justification" and "excuse", the legal prohibition of intentional killing outside self-defence, the incommensurability of goods, and the legal constrains to the use of lethal force set by normative positions: obligations, responsibility, rights, and authority. For each of these principles and constrains the possible application to the programming of autonomous vehicles is discussed. Based on the analysis, some practical suggestions are offered.

Keywords Ethics of autonomous vehicles · Ethics of self-driving cars · Robot ethics · Legal doctrine of necessity · Trolley problem

✉ Filippo Santoni de Sio
f.santonidesio@tudelft.nl

[1] Department Ethics/Philosophy of Technology, Delft University of Technology, Jaffalaan 5 - 2628BX, Delft, The Netherlands

1 Introduction

How should "Autonomous Vehicles"[1] (AVs) be programmed to behave in the event of an unavoidable accident in which the only choice open is one between the causing of different damages to different objects or persons?[2] Different fictional scenarios have been recently proposed in the philosophical literature (Lin 2015; Gerdes and Thornton 2015; Bonnefon et al. 2016). Establishing what kind of decision and behavior is morally permitted, prohibited, or obligatory in such emergency situations is a well-known hard philosophical problem.[3]

Bonnefon et al. (2016) have interpreted the issue of ethical programming of AVs in emergency mainly as one of public acceptance of new technologies, and have suggested that car manufacturers' decisions are supported by experimental ethics, that is empirical studies on lay people's feelings, opinions and intuitions on the moral acceptability/obligatoriness of different programming options. Gerdes and Thornton (2015) have instead taken the perspective of machine ethics and have made a first exploration of the possibility of implementing traditional ethical views into the programming of self-driving cars, for instance by programming machines to do utilitarian calculi to minimize the negative effects of their decisions while at the same time respecting some deontological constrains, for instance by always trying to avoid hitting human persons. However, as recognized by Bonnefon et al., and Gerdes and Thornton respectively, both lay people and philosophers disagree about what is morally prohibited, permissible or obligatory in scenarios where different fundamental interests and values are at stake, so that neither experimental ethics nor philosophical ethics seem at the moment able to offer car manufacturers and policy makers any clear indication for addressing this issue.

In order to make some steps forward, in this paper I suggest to take a *legal-philosophical* approach and to start taking a critical look at how the law has already

[1] By "Autonomous Vehicles" (AVs) I here mean civilian (as opposed to military) automated driving systems with "full automation", that is "full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*". In a full-automation vehicle the human driver cannot intervene in any way. See the SAE International's On-Road Automated Vehicle Standards Committee published the SAE Information Report: (J3016) (n.d.), "Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems". A summary of the report is available at http://www.sae.org/misc/pdfs/automated_driving.pdf AVs are often referred to as "self-driving cars".

[2] Admittedly, this is probably *not* the most urgent practical issue raised by the recent technological developments in automated transport. In fact, whether, when and according to which regulations AVs should be allowed on the roads are open normative questions themselves, so that reflecting and deciding on *these* issues is probably a more urgent practical task. However, assuming that AVs will at some point be allowed on the roads, that of regulating their programming in emergency situations will also become an urgent task. Moreover, reflecting on the ethical issues involved in the programming of AV may be one relevant element itself in the broader normative issue of the ethical acceptability of AVs. For a survey of the ethical issues involved in the introduction of AVs in society see Santoni de Sio (2016).

[3] In the road traffic scenarios discussed in this paper the subject actively harms someone in order to avoid damaging another. Therefore the closest reference in the philosophical literature is the discussion on the thought experiment known as the "trolley problem"(Foot 1967; Thomson 1985): There is a runaway, unstoppable, trolley. The trolley is headed straight for five people tied up and unable to move. A bystander is at some distance off in the train yard, next to a switch. If she throws the switch, the trolley will switch to a different track, where there is one person. Less relevant for this paper are rescue scenarios, like those discussed in the philosophical literature after the seminal paper by Taurek (1977). In those rescue scenarios the person should not directly damage anyone, but just choosing whom to benefit.

regulated similarly difficult choices in other emergency scenarios. In particular, I propose to consider the legal doctrine of necessity as it has been elaborated in the Anglo-American jurisprudence and case law. The doctrine of necessity seems a promising starting point as it regulates emergency cases in which human agents have *intentionally* caused damages to life and property in order to avoid other damages and losses, when avoiding all evils is deemed to be impossible.

The main methodological idea behind this approach is John L. Austin's (1956) suggestion that legal reasoning may be a sharp instrument of clarification of complicated philosophical questions. According to Austin, the reflections of lawyers – with their standing attention to real-life cases, their need to reach well-grounded and clear answers, their reflection partly independent from philosophical abstract theorizing and bias – may offer a fresh start to address difficult philosophical problems. Whereas I do believe that looking for fresh solutions to new or hard ethical problems is ultimately a philosophical enterprise, in line with Austin's suggestions I also think that philosophical reflection may sometimes benefit from considering legal principles and norms; in fact, legal principles are often the result of a combination of abstract moral principles and practical considerations deriving from the close observation and comparative analysis of real cases; moreover, legal norms are often an explicit attempt to cope with the fact of disagreement about general normative principles by finding a "reasonable compromise between principles and interests in contrast" (Hart 1961: 128).[4]

This paper has a twofold goal: to make a rational reconstruction of some major principles and norms embedded in the Anglo-American law and jurisprudence on necessity; and to start assessing which, if any, of these principles and norms can be utilized to find reasonable guidelines for solving the ethical issue of the regulation of the programming of AVs in emergency scenarios in which some serious damages to property and life is unavoidable.

The rest of the paper is arranged as follows: Section 3.1 presents the legal limits to permissible intentional killings outside self-defence as defined in two landmark legal cases; Section 3.2 presents the problem of incommensurability as grounding some key conceptual and moral reasons behind the legal skepticism towards what I call a simple utilitarian approach to killing under necessity; the successive sections present the legal constrains to the use of lethal force set by normative positions, namely: obligations (Section 4.1 and 4.2), responsibility (Section 4.3), rights (Section 4.4), and authority (Section 4.5). For each of these principles their possible application to the programming of AVs is discussed. In the next Section 2, the doctrine of necessity and the distinction between "justification" and "excuse" are introduced. The final section recapitulates the results of the paper.

2 Necessity, the Justification V. Excuse Debate, and AVs

The so-called doctrine of necessity concerns the legitimate exceptions to the compliance to a legal prohibitive norm. Behaviours that are prima facie prohibited by criminal law may be permitted under exceptional circumstances. Necessity is therefore a *defence*: a claim that a defendant may raise to prove that her behavior – for instance intentionally destroying private

[4] Peter Asaro (2011) endorses a similar approach with specific reference to robot ethics.

property – though generally prohibited by the criminal law, should not count as a crime under the circumstances – for instance if one does the destruction in order to escape from a fire.[5]

Lawyers usually draw a distinction between two kinds of defences: justifications and excuses (Fletcher 2000). The offender is *justified* when her action, though generally prohibited, is carried out in exceptional circumstances that eliminate the wrongness of action, as for instance in cases of killing in legitimate self-defence. The offender is *excused* when her wrong action was done under conditions that eliminate her culpability, as in cases of non-culpable ignorance of relevant circumstances: someone takes someone else's property in the reasonable belief that it is her own; or coercion: a bank clerk handing over the bank's money at gunpoint. Both justifications and excuses leave the offender off the hook, but whereas excuses operate on the agent's culpability without affecting the wrongness of her action; justifications make an otherwise wrong action lawful.

Some courts have interpreted necessity as an excuse: agents operating under necessity are not culpable insofar as their will is overborne, like in standard coercion cases, and therefore their actions have to be considered as "morally involuntary" (Fletcher 2000). Reading necessity as an excuse allows to explain why it typically applies to one-off "extraordinary and sudden" scenarios, in which agents are in panic or otherwise under exceptional psychological pressure, like in the *Herald of Free Enterprise* shipwreck, where some passengers removed by force and threw in the sea another passenger who was blocked on a rope ladder on the way to safety (Smith 1989); but it does not apply to cases like the fictional forced organ transplant by Thomson (1986:95) (McSherry 2002:14). However, as noted by Dennis (2009) notions of 'overborne will' and 'moral involuntariness' are hardly an accurate portrayal of other situations in which necessity typically applies like, for instance, that of "a doctor weighing up the merits of a particular medical procedure for a patient unable to consent"(35). Indeed, necessity has played a key role in the recent *Re A (conjoined twins)* (2001) case: Gracie and Rosie Attard were conjoined twins joined at the abdomen. The medical evidence indicated that Gracie was the stronger sibling who was sustaining the life of Rosie and that should the twins be surgically separated Gracie had a 94% survival rate while Rosie was guaranteed to die. If they were left conjoined then Gracie was predicted to die before they were six months old, causing also Rosie's death. The Court of Appeal authorized the surgical intervention, and one of the judges, Lord Justice Brooke, explicitly invoked necessity as a defence for causing the death of Rosie (*Re A* 2001).[6]

If necessity were only an excuse based on the weakness of human will and motivation, then it could not arguably apply to any programmed behaviour of AVs. If the point of necessity

[5] In order for any defence to apply, a prima facie legal offence must have been committed; and a behavior counts as an offense only if it is a human voluntary action as opposed to an involuntary reflex, a physically coerced behavior, and the like (Moore 2010). One may then wonder whether in order to apply the defence of necessity to the behaviour of AVs one needs to attribute these vehicles the capacity to commit an offense, that is to perform a legally assessable action. Not necessarily. One only needs to assume that some human agents – be it the programmer, the manufacturer, the owner – can legitimately held responsible for the behavior of the automated vehicle, and that this human agent may raise a defence of necessity when (legitimately) called to answer for the damages or losses caused by the intelligent pre-programmed behavior of a vehicle. Whether this assumption about the traceability of responsibility is reasonable in relation to all possible future scenarios involving autonomous agents is a controversial issue (see the so-called "responsibility gap" problem, as originally framed by Matthias (2004) and Sparrow (2007). However, in this paper the focus is on scenarios in which this assumption can be safely made, because a programmer has intentionally programmed the behavior of a vehicle in an emergency situation, based on the reasonable belief that the vehicle will behave accordingly, and this actually happens.

[6] More on this in section 3.1 below.

were to permit to do what is otherwise prohibited when under the emotional pressure of tragic and sudden circumstances, then it would always be impermissible to deliberately instruct in advance an artificial agent to damage anyone. However, this is not the case as the law already allows for some intentional damaging, even some killings done after measured exercises of judgement.

3 Legal Skepticism on the Simple Utilitarian Reading of the Doctrine of Necessity

Recent empirical studies suggest that many laypeople would not object to a program that makes AVs swerve and kill a bystander in order to save more people from a potentially lethal accident (Bonnefon et al. 2016). And some utilitarian philosophers think this would also be the right thing to do from an ethical point of view.[7] The reasoning goes as follows: 1. in the presence of a tragic unavoidable choice between two evils, one ought to choose the lesser evil; 2. the lesser evil between intentionally causing the death of one and intentionally causing the death of more than one is intentionally causing the death of one; 3. When confronted with an unavoidable choice between intentionally causing the death of one and intentionally causing the death of more than one, one ought to (and, *a fortiori*, can) intentionally cause the death of one person rather than that of more than one. I call this the simple utilitarian reading of the doctrine of necessity.

In the next section I will present two landmark legal cases which have set some strict limits to the permissibility of intentional killing under necessity; in the successive two sections I will make a rational reconstruction of two sets of reasons grounding the legal skepticism towards a simple utilitarian reading of the defense of necessity, that is: the problem of incommensurability, and the importance of the constrains to the use of force set by normative positions: obligations, responsibility, rights, and authority. For each of these the application to the programming of AVs is discussed.

3.1 The Prohibition of the Intentional Killing of the Innocent in the Law

Until recently, the English law has taken a restrictive approach to the interpretation of the doctrine of necessity where human lives are at stake, by prohibiting the intentional killing of a person who is not posing any direct threat to the life of the defendant, and so is not a legitimate target of self-defence force.

In *R v Dudley and Stephens* (1884), the English law landmark case, the seamen Tom Dudley and Edwin Stephens were shipwrecked along with two other men. Thinking that they would be soon all starving, Dudley and Stephens killed for food one of the other men – who they have reasons to think was about to die anyways. Rescued and returned home, they were convicted of murder, based on the principle that no innocent life should be taken under *any* circumstance.[8]

[7] See the debate on the so-called trolley problem referred to in note 3 above.

[8] The verdict contained a recommendation for mercy, which was eventually accepted by the queen: the death sentence was turned into a six months' imprisonment.

However, things have changed in 2001, with the *Re A (conjoined twins)*, where a surgical intervention certainly causing the death of one of the conjoined twin was done in order to save the other twin.[9] As innovative as it might have been for English law, *Re A* still allowed for intentional killing under necessity under very specific circumstances, namely: a) the person killed was here the one (involuntarily) impeding the survival of the other[10]; b) the person killed would have certainly died anyways in a short time; c) the killing has been committed with the official permission of a public authority (i.e. a judge), after careful assessment of the relevant facts and the opinion of experts.[11]

3.1.1 AVs and the Intentional Killing of the Innocent

Leaving aside the role played by a public authority in the permission of the killing in Re A, which will be discussed in a separate section,[12] it is at least in principle possible to imagine emergency scenarios involving AVs with features analogous to *Re A*. Consider the following one. A fatal road accident is about to occur, in which vehicles A, B, C, D, E, and F will be involved and their passengers will all die because the collision will eventually make all vehicles fall from a high bridge over a dry rocky canyon; while there is no way for the AV A to avoid being involved in the accident, vehicle A may be programmed to pick which vehicle to hit among those already doomed to be involved in the collision and to fall from the bridge, and by which angle at which speed; the AV A clearly detects that the non-AV B is out of control as its driver has been pushed out from the vehicle and the only passenger onboard is a toddler on the backseat; vehicle B plays a crucial role in the imminent death of all actors involved in the accident, because, according to two different possible variations, either it is the one that will cause the violent collision which will make all vehicles fall from the bridge, or it is the one that will cause all the vehicles to fall from the bridge after the collision; in both variations the only way to avoid that all the vehicles A B, C, D, E and F are pushed down the high bridge, is for the vehicle A to target vehicle B at such an angle and a speed that B, and only B, will be pushed down the bridge while all other vehicles and their passengers will be saved. By anticipating a scenario in which all these conditions apply, Aisha, the programmer of the AV A, had decided to program the vehicle in such a way that, finding itself in such a scenario, vehicle A does exactly what is required to hit vehicle B down the bridge, thus causing the death of the vehicle B's baby passenger while saving the life of all the passengers of vehicles A, C, D, E, and F. It may be argued that Aisha should be acquitted from the murder of vehicle B's passengers, even though she has intentionally programmed her vehicle A to hit the vehicle B, directly causing the death of its passenger: she seems to be in a similar position of the doctors in Re A, who was allowed to kill one of two babies in danger to avoid the death of both of them.

[9] See section 3 above

[10] One of the judges of the Court of Appeal, Lord Justice Alan Ward, even tried to frame the *Re A* case as one of self-defence.

[11] It is arguable whether *Re A* is comparable to *Dudley and Stephens* as the epistemic position of the actors may be relevantly different and the killed one wasn't causing the death of the others. However, Re A is arguably similar to other cases presented in the literature on necessity: for instance, Dennis (2009) remarks that After 9/11 some eminent lawyers have suggested that necessity may allow the shooting down a hijacked aircraft, inevitably killing the passengers and crew, before it will crash into a tower block killing thousands more. Christie (1999) had already pointed out that necessity may apply to killing done by seamen, mine superintendents or mountaineers when the three above-mentioned conditions realize.

[12] Section 4.5 below.

Moreover, Aisha might even insist that her position is morally safer than that of the doctors in *Re A*, as she took her decision on hypothetical and unknown future scenarios as opposed to a real and known one, and so whereas she did the programming by relying on some general features of those scenarios, she could not know all the details of the emergency scenario in advance, for instance that a baby boy as opposed to a four-year old child as opposed to no one would be sitting on vehicle B at the moment of the crash. So, if one thinks, with Christie (1999), that: a) what is morally problematic about the simple utilitarian reasoning in trolley-like scenario[13] is that the act of killing is intentional, and b) "intentional" means deliberately directed at one or more individual persons; then, a programmer like Aisha might appeal to the doctrine of double effect[14] and claim that she has not intentionally killed anyone even though she has clearly anticipated that her behavior would cause someone to die. Be that as it may, the doctrine of double effect would arguably save Aisha only from the accusation of intentional murder, not from that of manslaughter, or even that of murder by *dolus eventualis*.[15]

Admittedly, this is a far-fetched scenario, which does not represent the current capacities of sensing of autonomous vehicles: cars cannot and will arguably not able any time soon to make the kind of sensing and computation described in the scenario above. As it is in this kind of scenario that the current legal approach would allow for a justification of necessity, based on a first review of current case law on necessity, we seem to be left with the idea that AVs manufacturers would not easily be (legally) justified in intentionally (and systematically) targeting any (group) of bystanders in order to prevent another damage or injury that the vehicle is about to cause.

Many philosophers and laypeople find this conclusion dissatisfactory and think that the widespread reluctance to accept the introduction in the law of a simple utilitarian reasoning about emergency killings is ultimately irrational. Harris (1975) famously pointed out that even accepting the idea that the goal of morality and the law is that of protecting some basic goods – typically but not only: human life – it seems that when it is impossible to protect all the instantiations of that good – everyone's life – we should be at least allowed, if not obliged, to try to save as many lives as possible. Nonetheless, even though necessity has indeed sometimes been presented as a "choice of evil" defence in the US law[16] and this interpretation has been defended as correct by some eminent legal scholars (Robinson 1975), a simple utilitarian reading of the "lesser evil" principle in cases of killing remains rather unpopular among lawyers.[17] Understanding the reasons behind the lawyers' skepticism may help clarifying some conceptual and ethical reasons against a simple utilitarian programming of AVs.

[13] See note 3 above.

[14] See for instance Quinn (1989).

[15] Interestingly, in the *Re A* case one of the judges, Lord Justice Robert Walker, did deem the intervention as permissible by relying on an interpretation of the surgeons' intentions: "Mary's death would not be the purpose or intention of the surgery, and she would die because tragically her body, on its own, is not and never has been viable".

[16] "Conduct that the actor believes to be necessary to avoid a harm or evil to himself or to another is justifiable, provided that … the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged" (MPC 3.02) The official commentary makes it clear that section 3.02 of the Model Penal Code was not meant to preclude the raising of the lesser evil defense in a case of intentional homicide.

[17] Even the Model Penal Code (1958: § 3.02 cmt. 3), despite its declared acceptance of the "choice of evil" principle, still remains silent as to whether the defense is available in cases of homicide.

3.2 Incommensurability

One important reason behind the lawyers' skepticism towards simple utilitarian calculi in emergency scenario is the problem of incommensurability. This and the next section are devoted, respectively, to a rational reconstruction of some variations of the problem of incommensurability as they have been presented in the legal literature on necessity, and to an answer to the question as to which if any of these variations may constitute a good reason against the application of a simple utilitarian reasoning to the programming of AVs.

A first problem with a simple utilitarian reading of necessity is constituted by the difficulty of making a reasonable measuring of the values of different goods at stake in an emergency scenario. Making a balance of the values of different goods may be difficult even when only material goods not human lives are at stake. According to Christie (1999: 1000) someone's property may not be destroyed to save one's own property or someone else's property, because it is often difficult to establish an objective standard for the evaluation of the value of different objects at stake; according to his example, it is impossible to establish whether it is more valuable a van Gogh's painting or an old Bible handed down across many generations in a religious family. Moreover, there may be relevant differences in the "comparative ability of the sufferers to sustain the loss" of a similar good (*Latta v. New Orleans & N.W. Ry* (1912:254).[18] One might point to similar considerations also, and arguably even more, in relation to cases when killing one person is necessary in order to kill another or more persons[19]: given that all lives are different, their value is not measurable based on objective standards, and their loss usually affect other persons, it is impossible to establish a priori that, for instance, (any) five lives are more valuable than (any) one life.

A second problem with the simple utilitarian approach to emergency scenarios is constituted by the difficulty of assessing the weight of long-term and less visible harms caused by allowing for non-compliance to a general norm prohibiting the intentional damaging of basic interests like life, physical integrity and property. Dennis (2009) points to the difficulty of measuring the moral and societal harms embedded in the intentional violation of a legal norm – in terms of bad example, long-term loss of credibility of the law as well as the wrongness of acting against a public good – and to compare these with the material benefit that one particular individual may achieve through the violation of that norm on one specific occasion, for instance saving some lives. Dennis' remark reveals, I think, a more general issue: unlike philosophical thought experiments and laypeople occasional moral judgements, the law should not only consider the desirability of the consequences of different course of actions on a single occasion, but it should also reflect on the long-term societal effects of rules and policies allowing for or even encouraging certain kinds of behaviors. It may be difficult to factor these long-term effects in a simple utilitarian calculus.

[18] See also the slippery slope and Pandora box concerns in section 4.5 on Authority below.

[19] This problem is particularly relevant when the *active* killing of one (innocent) person is necessary to save one or more other persons (see e.g. Dudley & Stephens). Given that lives are not comparable goods, nobody should be allowed to intentionally kill someone on the basis of the assumption that her one life is less worth of one or more other lives. On the other hand, the argument is less relevant in rescue scenarios (Taurek 1977), where no active killing is at stake and the choice is between two alternative rescue operations aimed to save different (groups of) persons. Here incommensurability may certainly speaks against the obligation to save the higher number, but it does not speak against allowing for a choice. In fact, even if it is not possible to decide which intervention is morally better, it is clear than any rescuing intervention is better than no intervention.

To be clear, the problem of incommensurability may be read in three different ways: as a conceptual, an epistemic, or a moral problem. According to the conceptual reading, the problem is that it is impossible to compare the value of goods like different lives: as this value is determined by subjective evaluations there are no objective standards or metrics for making reliable measurements. According to the epistemic reading, whereas there may be in principle reliable objective standards to decide whether one particular life is more valuable than other particular lives, agents involved in dilemmatic circumstances may not have the information, the time or the capacity to make a proper evaluation. Moreover, they may not be able to assess the long-term consequences of their actions. According to the normative reading, no matter whether lives and property can be compared or their values measured, the rights to life and property owed to different individuals cannot be compared or traded with other rights: they just have to be respected. Those discussed in this section are mainly conceptual and epistemic reasons: the right-based reasons will be discussed in a separate section below.[20]

It is worth noting that from a conceptual and epistemic point of view, the incommensurability problem usually does not constitute an obstacle to utilitarian calculi in circumstances where some property should be taken or damaged in order to save one or more lives, as any property may be arguably be considered as less valuable than any life. In *Ruiz v Forman* (1974), for instance, a lorry driver was acquitted from trespassing and damaging the claimant's property, as this was done in order to avoid the collision with an upcoming vehicle.[21]

Moreover, it has been claimed that as serious as it may be in most emergency circumstances, the conceptual incommensurability argument does not apply to catastrophic scenarios where the sacrifice of the life of one or few persons may prevent the death of a huge number. According to Michael S. Moore, "*A* may not torture *B* to save the lives of two others, but he may do so to save a thousand lives" (Alexander and Moore 2015, referring to Moore 1997, ch. 17). However, as pointed out by Christie (1999), a right-based approach may lead to a different conclusion, and in fact human right international law adopts a stricter approach, one that prohibits torture under *any* circumstances.[22]

3.2.1 Incommensurability and AVs

Is incommensurability a problem for an utilitarian programming of AVs? If one sees incommensurability of lives and properties as a mere epistemic rather than a conceptual or moral problem, then it can be argued that the artificial intelligence of AV may in principle be able to eventually overcome this problem. An evaluation of factors like age, profession, health and financial conditions, family status and social relations of the persons involved in a potential accident may be in principle included in the programming of a vehicle, much in the same way in which these factors are included in the insurance companies' assessments of the availability and costs of insurances for different clients.[23] Even more so, given that AVs may in principle outperform humans

[20] More on rights in Section 4.4 below.

[21] Christie (1999:994–5) even claims that no compensation should be due to owners of the damaged property, in order to avoid the risk of disincentivizing the causing of the material damages necessary to avoid killings with the fear of future compensation.

[22] See Section 4.4 below.

[23] This is, again, a unrealistic scenario given the current status of technology, but it is not conceptually impossible.

in their power of calculus, especially under the pressure of time and/or emotionally challenging circumstances.[24] However, the assessment of long-term or cumulative (negative) effects of a policy remain a problem even in a scenario of superintelligent machines. These effects may include a negative impact on people perception of safety or a negative impact on the life of people systematically disadvantaged by the utilitarian policy.

Moreover, if one sees incommensurability rather as a conceptual or a moral problem, no increase in the acquisition of information or in the power of calculus by artificial systems may overcome it. From these perspectives, the problem is either that there are no objective facts to compare and/or there is no metrics to compare these facts, or that the individual right to property and life does not depend on the objective empirical features of different individual goods or persons.[25]

As for the cases where the conceptual reading of incommensurability arguably does not apply: the destruction of property to save lives and the big catastrophes scenarios, these may be ones in which AVs may indeed be programmed to follow a simple "utilitarian" procedure. So, it may be permissible or even obligatory to intentionally damage private property in order to save lives, like the human driver did in *Ruiz v. Forman*. Cases in which the killing of a thousand persons is at stake are more difficult to imagine in relation to civilian AVs.[26]

In summary, the existing law poses strict limitations to the performance of simple utilitarian calculi to justify the intentional damaging of basic interests like property and human life in emergency scenarios. This attitudes has been justified in the literature with reference to the problem of incommensurability. As most of the problems of incommensurability remain valid in the case of AVs, this may be arguably seen as one reason for adopting a skeptical attitude towards a simple utilitarian approach also in this new case.[27]

4 Normative Positions: Obligations, Responsibilities, Rights, and Authority

Some philosophers have tried to address the permissibility of killing under necessity by reflecting on thought experiments like the trolley problem.[28] This approach has been criticized, among other things, for neglecting the potential relevance of the normative positions – in particular the responsibilities – of the various agents involved in the situation (Wood 2011). With this in mind, in order to better understand the moral foundation of the legal doctrine of necessity and to carve the limits of its application, in the next sections I will turn to the analysis of the role of normative positions of different agents involved in life-threatening emergency scenarios; for each normative position I will present some insights from the legal literature, and I will try to assess their relevance for the case of AVs.

[24] Such a scenario would involve serious privacy issues, but this is a different problem.
[25] The moral reading of incommensurability will be also discussed in the Section 4.4 on rights below.
[26] Though they may be thinkable in relation to military robots.
[27] Additional reasons are presented in the following sections.
[28] See note 3 above. Recently, Nyholm and Smids (2016) have made a helpful critical assessment of the idea of the AVs programming being an "applied trolley problem".

4.1 Contractual Obligations

Firstly, in defining the responsibilities of agents involved in emergency scenarios, the law has stressed the importance of contractual obligations. In the landmark case *United States v. Holmes* (1842)[29] some ship crew were tried for throwing 14 passengers in the sea in order to prevent the sink of a lifeboat on which they were piled with many others, following a shipwreck. Justice Baldwin explained to the jury that in assessing the sailors culpability for the killing of the passengers, they should be reminded that a shipwreck is not a scenario in which people "owe no duty to one another": there were specific normative relations between people involved.[30] In particular, sailors owed certain duties to the passengers, and breached those duties by jettisoning some of them. Justice Baldwin explains that there are "obligations which rest upon the owners of stages, steamboats, and other vehicles of transportation.": "in consideration of the payment of fare" and "Having, in all emergencies, the conduct of the journey, and the control of the passengers, the owners rest under every obligation for care, skill, and general capacity; and if, from defect of any of these requisites, grievous injury is done to the passenger, the persons employed are liable" (Cohan 2006:150-151). This relation cannot "be changed when the ship is lost by tempest". Therefore, whereas the captain and a sufficient number of seamen must be preserved to navigate the boat, "supernumerary sailors have no right, for their safety, to sacrifice the passengers. The sailors and passengers, in fact, cannot be regarded as in equal positions. The sailor . . . owes more benevolence to another than to himself." (151)

4.1.1 Contractual (and Extra-Contractual) Obligations and AVs

A case with some similarities with *Holmes* would be one in which the AV is a public service like a bus. Here the owner of the bus service and her employees arguably have a duty of care towards the passengers similar to the one owed to the passengers of a ship by the ship owner and his employees described in *Holmes*. However, here the scenario is complicated by the presence of two additional groups of agents: the company producing the vehicle, and the potential victims of a road accidents. The reasoning of Justice Baldwin in Holmes does not help define these additional relations, as it concerns a scenario in which the owner of the means and his employees have "in all emergencies, [a] the conduct of the journey, and [b] the control of the passengers", but this might not be the case with automated vehicles: the owner of an automated bus does not have either the "conduct of the journey" or the control of the passengers; the vehicle manufacturers arguably have (more of) it; in addition, all the people involved in Holmes' shipwreck are either members of the crew or passengers, whereas road accidents may typically involve third parties who are not bound by specific contractual relations to the vehicle's owners.

As for the control issue, even though the owner of a public service does not directly control the programming of the vehicles he deploys, he arguably has an obligation to maintain some form of control or supervision over the programming of her vehicles, by making sure that she

[29] This is the US law landmark case on murder under necessity: eight seamen, including Holmes and thirty-two passengers, got into a lifeboat, following the ship collision with an iceberg. The lifeboat was filled beyond capacity and started to leak upon being launched; the crew thought that the boat was too unmanageable to be saved. In order to save the boat and at least some of the passengers, Holmes and the rest of the crew threw over 14 male passengers to save the boat. Upon their return to the United States, a grand jury indicted Holmes for manslaughter committed upon the high seas.

[30] I rely here on the discussion of the case in Cohan (2006)

does not deploy a vehicle whose programming is in contrast with her duty of care towards her passengers. So, for instance, she may not be allowed to deploy a system designed to protect the vehicle and/or the company's employees on the vehicle at the expenses of the passengers.

As for the normative positions of third parties involved in an accident, even though the owner of a vehicle does not have any contractual obligation towards them, she certainly has an extra-contractual obligation to not damage them.[31] So, it seems that looking at the contractual obligations is not enough to solve the issue of what should be the programming of an automated public service bus in an event of an emergency scenario in which a choice is only open between exposing the passengers or rather some third parties to the risk of serious injury or death. This last point is arguably relevant also to assess the duties and responsibilities of private autonomous vehicles manufacturers. Certainly, car manufacturers are bounded to a contractual obligation with their clients and have a stringent obligation to protect their safety in the design of their vehicles; in this respect, they are in a position not too dissimilar than that of sailors towards their passengers. However, the case of autonomous vehicles differs from the ship case in one important respect: whereas in the boat case sailors can and arguably should sacrificing *themselves* for protecting the life of their passengers, in the autonomous vehicle case, car manufacturers may sacrifice *third parties* to protect their clients life. These other people's rights should be considered too.[32]

4.2 Conventional Obligations

A second kind of normative position potentially relevant to define the limits of the permissibility of killing in emergency scenarios is that deriving from conventional obligations. The philosopher Judith Jarvis Thomson (1990) has suggested that the killing of one or more persons in order to save another, bigger group of persons, for instance in a standard trolley-problem scenario,[33] may be permitted if the following additional circumstances realize: the helpless men trapped on the railway are all workmen and part of the same crew whose tasks are randomly assigned on any given day; when workers join the work crew, it is explained to them that their occupation is a dangerous one in which death or serious injury is a distinct possibility and that should an emergency situation arise in which the certainty of the death of a larger number of men can be averted by killing a lesser number, then this will be done. Thomson claims that reasonable workmen would enter such an agreement and so they may be killed should the tragic choice presents itself.

But Christie (1999:1017) wonders what the relevance of Thomson's claim would be if, as a matter of fact, in the real world no one ever enters into these kinds of arrangements, for instance by signing a contract. Moreover, it is very dubious whether such contracts would be even legally valid. In fact, as the victims' consent cannot per se justify the commission of a crime like a murder (the maxim *volenti non fit iniuria* does not apply to criminal assaults), either the killing of the lesser group is justified on independent grounds, for instance, as being done according to a just legal procedure, as suggested by Thomson,[34] or it is not. If it is justified, than the killing may be justified and the consent of the victim is not necessary, if the killing is not justified then the killing counts as a criminal assault, and the consent of the victim is immaterial.

[31] See sections on responsibility and rights below, section 4.4.
[32] I will discuss this in the section 4.4 on rights below.
[33] See note 3 above.
[34] More on this in the section 4.5 on authority.

In addition, whereas at least tort law does allow for *volenti non fit iniuria* contracts – agreements in which by acknowledging and accepting the risks involved in a given activity agents waive their opportunity to make a claim against the other party in the event of a damage – these contracts are subjected to stringent conditions of full knowledge and freedom of consent. Assuming that by being a professional in that domain a worker has a full understanding of the risk he is entering by joining a certain work team, it remains highly debatable whether this consent should be considered as free. In fact, the law has put increasing restrictions on the application of a *volenti* contract on the workplace, due to the risks of coercive agreements for the employees engaged in risky activities.

Quite interestingly, in the *Holmes* shipwreck case described above both the prosecution and the defendant's attorney insisted on the relevance of "the customs of the sea's" prescription to cast lots in exigent circumstances. Justice Baldwin even instructed the jury that had lots been used to select the victims,[35] the defence of necessity might have been available to the crew who jettison some passengers in order to avoid the sinking of a lifeboat (Cohan 2006:154). So some customary rules and conventions may matter, after all.

However, Cohan states that it is an "unsettled issue" whether Justice Baldwin's dictum on drawing of lots was an accurate summary of the law at the time of *Holmes* or today. Moreover, like Christie did about Thomson reasonable agreement between workers, Cohan also points to some difficulties in giving precise content to such conventions and enforcing them: Should everyone consent to the drawing of lots in order for the drawing to be binding for everyone? Should people be forced to participate in the ballot? If someone refused to participate, could that person's lot be selected by proxy?

4.2.1 Contracts, Conventions and AVs

If the appeal to existing conventions or tacit agreements to decide how to regulate killings in emergency scenarios in dangerous long-established activities like sailing or mining may sound morally problematic, this appeal is probably utterly useless in the case of a new practice like fully automated transport, where no conventions are yet in place. And even assuming that there are shared conventions among car manufacturers' companies about which courses of action to program in an event of an emergency situation (for instance: always protect your client), these may arguably not replace long-established ethical and legal principles in areas where life and physical integrity of people is directly at stake. Moreover, as already mentioned, the presence of third parties not bounded by any contractual or conventional relationship to the car manufacturers make the car case significantly different than the boat one.

Certainly, the idea of solving some emerging issues in the regulation of the design, production and license of autonomous vehicles by establishing in advance contracts and agreements between potential parties involved in future accidents may indeed sound as a more reasonable solution than simply letting car manufacturers free to decide. In particular, high level of autonomy will make it envisageable and recommendable the creation of new forms of insurances and possibly new forms of civil responsibility, in addition to the existing product liability and fault responsibility.[36] However, areas that have a public interest like the regulation of *intentional* behavior in road traffic, where basic rights like life and physical integrity are at

[35] Also the different statuses of passengers and crew members are mentioned in this passage; more on this in the section 4.3 on responsibility.

[36] See Pagallo (2013), Smith (2015).

stake are usually not covered by private contracts but by State regulation and the (criminal) law. So it seems quite unlikely that professional conventions or private insurance agreements may determine the regulation of programming of AVs in emergency scenarios where the (intentional) damaging of human lives is at stake. One possible exception might be the signing of a contract by the user of AV in which he acknowledges and accepts that the vehicle is programmed in such a way as to sacrifice its passengers – for instance by directing the vehicle against a wall – in order to avoid hitting other vehicles. This might be legally acceptable, provided a well informed consent were in place,[37] as it would potentially affects only the users of the vehicle who signs the contract. For this same reason it is doubtful whether anyone would ever freely sign such an agreement.[38]

4.3 Responsibility

Responsibility may also affect the normative positions of persons involved in life-threatening scenarios like road accidents. In fact, according to the legal doctrine of contributory negligence, a person who has been damaged in an accident also due to her own fault may be assigned a lesser compensation for the damage suffered. For instance, In *Goddard and Walker v Greenwood* (2003), a pedestrian hit by a car was found liable for 80% of his injuries for crossing with the lights against them. However, even in tort law contributory negligence of the claimant rarely completely eliminates the car driver's liability. In the same *Goddard & Walker v Greenwood* case, the Court of Appeal explained that the driver still had to cover some of the damages, because the fact that the lights were green in his favour does not obviate the need to keep a close look out. And In *Eagle v Chambers* (2004), LJ Hale explicitly said that even if the claimant was struck whilst dangerously walking along a dual carriageway, the car driver should be assigned a higher share of liability. Such a high burden imposed on car drivers – the judge explained – reflects the fact that cars are "potentially dangerous weapons".

More importantly, it is a well-established legal principle that contributory negligence is not a defense in a criminal prosecution for *negligent homicide* by a motor vehicle unless negligence on the part of the decedent is found to be the sole proximate cause of the death (*State v. Scribner* 2002: 741), that is, roughly, unless the behavior of the victim entirely explains the accident. Therefore, from a criminal law perspective as far as the accident occurred *also* due to a deliberately or negligently dangerous behavior of the driver, this is criminally liable of manslaughter or murder – no matter whether the victim was drunk or negligent or otherwise contributed to his own death, "for in this consists a great distinction between civil and criminal proceedings" (*R v Swindall and Osborne* 1846).[39]

4.3.1 Responsibility and AVs

Current law puts a high duty of care on car drivers, based on the fact that they handle "potentially dangerous weapons". It is reasonable to think that a similar or even higher burden should be put on companies that produce and commercialise AVs. In fact, even assuming that future AVs will be safer than current ones in the sense of causing lesser accidents due to the elimination of the impact of human drivers errors, AVs will still be potentially more dangerous

[37] Full information is a key requirement for a *volenti non fit iniuria* contract to be valid.
[38] See Bonnefon et al. (2016) More on this in the section 4.3 on Responsibility.
[39] Similarly in *R v Longbottom* (1849).

than bicycles or pedestrians in the sense of having a higher potential for causing serious damages to third parties in the event of a crash. In addition, according to current tort law the duty of care of drivers towards pedestrians extends as far as to cover most of the damages that could be prevented by the driver's diligent behavior, no matter how negligent the behavior of pedestrians might be. Therefore, the enhanced ability for crash avoidance brought by artificial intelligence may put an even higher duty of care on the shoulders of autonomous vehicles manufacturers.

Finally, from a criminal law perspective, drivers – and so, arguably, future producers of AVs – have a stringent duty to not intentionally or negligently damage third parties, and in particular weaker parties like pedestrians or cyclists, again no matter how negligent the third parties' behavior may be. This seems to make unacceptable from the start any suggestion about programming AV to preferably target, in an emergency scenario, some third parties who are somehow responsible for their being in a dangerous position, for instance pedestrians crossing with red lights, cyclists outside the bike lane, or bikers without helmet.

One may then wonder whether it would be consistent with the company's duty of care their programming the vehicle so as to preferably hit: a) other vehicles as opposed to pedestrians or cyclists, and b) vehicles with higher safety protections as opposed to less safe ones, in the attempt to reduce the number and seriousness of casualties (Lin 2015). Whereas the preference for hitting vehicles as opposed to pedestrians or cyclists may be justified by the asymmetry in the duty of care of car drivers/manufaturers and pedestrians/cyclists, the idea of intentionally targeting safer vehicles looks more problematic. In fact, from a normative point of view, all private vehicles' passengers have the same responsibility - i.e. that deriving from taking the risk of using private motor vehicles – and thus there seems to be no normative ground for intentionally penalising the passenger of a safer vehicle.[40] Of course, one may invoke common-sense or utilitarian considerations to defend such a programming. But it may be pointed out that when translated into a systematic policy, this may be self-defeating even from an utilitarian point of view, because it may end up discouraging the investments and use of safe vehicles, and so contributing to make the road more dangerous and the number of victims of car accidents higher in the long run.[41]

Finally, one may also wonder whether the obligation to try to avoid hitting pedestrians and cyclists is so stringent as to require the killing of the passengers of the AV when this is the only open alternative option. It is probably the case. Again, one of the points of the legislation on vehicular traffic is the protection from potentially fatal accidents, and the obligation to avoid the intentional or negligent collision with pedestrians and cyclists is an essential part of this legislation. A norm providing car manufacturers with a "license" to intentionally hit an innocent pedestrian (in order to save the vehicle passenger) would be in contrast with this basic principle. This certainly does not mean that car manufacturers should be allowed or even incentivized to program their vehicles to sometimes sacrifice their passengers, as this would be in contrast with the companies' competing duty to protect the life of their clients. But it certainly means that car manufacturers and policy-makers should be pressured to find ways to design vehicles, infrastructures and the legislative settings of road traffic in such a way as to avoid *both* the killings of pedestrians and cyclists *and* that of AVs passengers; for instance by

[40] If anything, one may argue that users of safer vehicles should be rewarded not penalised. The idea of flipping a coin to decide morally difficult *rescue* scenarios (Taurek 1977), where "doing nothing" is uncontroversially the worst option, is less appealing here.

[41] See also section 3.2.1 above. In the philosophical literature, a somehow similar objection was moved by Singer (1977) to Harris (1975) in relation to the "survival lottery" thought experiment.

designing infrastructures and norms in such a way as to make the realization of such a tragic dilemmatic scenario highly unlikely.[42]

4.4 Rights

The reluctance of lawyers to accept a simple utilitarian reading of the defence of necessity is also grounded in the protection of the right to life. From the point of view of criminal law, the importance of the right to life can be highlighted by looking at one specific feature of the justifications of necessity and self-defence. Remember the justification/excuse distinction[43]: Whereas excused actions remain forbidden and may thus be carried out exclusively by the agents acting under the excusing conditions, justifications like necessity and self-defence make generally prohibited actions lawful, so that people acting under justificatory circumstances may be also assisted by third parties. It is for instance lawful for a safe bystander to help someone damaging some private property to let them escape from a fire (necessity), and so it is for a safe bystander to intervene to help the victim of a lethal unlawful aggression to incapacitate and possibly kill her aggressor (self-defence); however, the aggressor cannot in turn invoke a self-defence justification for using violence against the rescuer, as the rescuer's violence is legally justified. Christie (1999) points out that this normative implication of justifications generates the following implication for a simple utilitarian readings of the doctrine of necessity. If, for instance in a trolley problem scenario,[44] one considers throwing the switch as permissible in the sense of justified as opposed to just excused, then from a legal point of view they should also accept the consequence that, for instance, friends or family of the potential victim may *not* try to stop the switch-throwing, and if they try they may be even legitimately prevented by force by the would-be switch-thrower, or by any other bystander.[45] By pushing the simple utilitarian reading to its extreme legal consequences, Christie wants to make vividly represented the innocent bystander's *right* to life – i.e. the right not to be killed - and its legal superiority over the still legitimate *request to be rescued* from the people in danger.

4.4.1 Rights and AVs

A reference to the rights of potential victims of automated vehicles accidents may complement and strengthen the conclusions of the previous sections on the prohibition of intentional killing and on responsibility. Not only pedestrians and cyclists are owed a special duty of care by motor vehicles manufacturers based on the responsibility of the latter for taking the risk of producing a dangerous machine; they also are owed a duty not to be killed based on their basic rights as human persons. This means that pedestrians and cyclists who would not be involved in any accidents but for the decision of the AV (manufacturer), may arguably not be involved in the accident by design, even though by doing so some lives of people already involved in the accident may be saved.

[42] See Santoni de Sio (2016) for a similar suggestion based on the Value-sensitive design approach.
[43] See section 2 above.
[44] See note 3 above.
[45] In the same way in which the parents of the conjoined twins of the Re A case could not stop the doctor performing the intervention that eventually save one of the two by killing the other.

4.5 Authority

Finally, an important feature of the rare scenarios where lawyers have been willing to recognize a necessity justification for the intentional killing of an innocent to save other innocent lives is the presence of a public procedure permitting or even commanding the final death-causing decision.[46] In the *Re A* (conjoined twins) case presented above, the doctors acted under the order of a tribunal when they performed the surgery which would cause the death of one of the twins to save the other.

The intervention of a public authority may be requested to avoid one or more of the following risks. First, even assuming that clear guidelines for justified killing were available, leaving the application of these in the hands of private citizens would open the door to serious risks of abuse; secondly, and relatedly, there are slippery slope or "Pandora box" risks, namely risks that once private citizens are allowed to make these reasoning in safe cases, they may start making these choices in less clear-cut cases (Norrie 2014). Thirdly, a public procedure may guarantee a more reliable assessment of the certainty of the future outcomes. Fourthly, a public procedure leading to a final authoritative command may sometimes be the only reasonable way to overcome the moral disagreement between citizens and to allow for a legitimate political and legal decision in ethically controversial cases.

4.5.1 Authority and AVs

It may be argued that all of the concerns above – risks of abuses, necessity of an official check on the certainty of future outcome, necessity of public procedure and authoritative decisions to solve hard ethical cases with a public relevance – also apply to the programming of AVs.

The public control over the management of emergency scenarios involving AVs might take two different forms. A first, simpler solution would consist of just giving vehicles manufacturers binding legal guidelines for the programming of their vehicles in emergency situations. A second, more sophisticated solution would be to also create a centralized (automated) system of coordination and monitoring of automated vehicle traffic, under the ultimate control of a public agency, which would be able to directly manage emergency situations according to the law (this might in turn involve the use of Artificial Intelligence systems of control).

In both cases the further question would remain as to which public authority should take the decision. It has been argued that in compliance with the principles of democracy, a democratically elected representative body rather than a judge should determine the hierarchy of values to be pursued in emergency scenarios, at least when life is not at stake (Dennis 2009). This seems to suggest that the case of AVs, where life is at stake, might be handled via the judicial application of the existing law. However, it can also be argued that scenarios opened by new technologies are so new that judges, not having precedent cases to rely on, may take idiosyncratic or otherwise arbitrary decisions about them, so that a prior intervention of a legislative body may be preferable.

5 Conclusions

Based on a critical reconstruction of some major legal principles and norms currently embedded in the legal doctrine of necessity, the paper has identified the following, tentative

[46] The importance of a public authority for deciding issues of necessity is stressed by Christie (1999).

normative principles for the programming of AVs in emergency scenarios where the damaging of some property or life is unavoidable:

1. As necessity works also as a justification not only as an excuse for deliberate damages, there might be in principle circumstances in which a vehicle may be programmed to kill (*Re A conjoined twins*);
2. Given the strong restrictions to the intentional killing of innocents outside self-defence (*Dudley and Stephens*), the problem of incommensurability of values and the right of life of persons, a program that allows for a vehicle to systematically hit persons who wouldn't be involved in the accident but for the vehicle decision seems unacceptable;
3. Based on the current legal constraints on killing under necessity (*Re A*), the intentional programming of the a AV to target another AV might in principle be permitted under very specific and complex circumstances, which seem very unlikely to realize in the next future;
4. The intentional programming of a vehicle to damage property (included the vehicle itself) to avoid damaging persons may be allowed and even made obligatory (*Ruiz v.Forman*);
5. Vehicles deployed as a public service maybe required to be programmed to not expose passengers to damages in order to protect the public service controller onboard, based on the duty of care of service providers to their passengers (*Holmes*); more generally, car manufacturers may claim to have a contractual obligation to protect their "passengers", but this should not be done at the expenses of the physical integrity of other road users, as this would be in contrast with car manufacturers' extra-contractual obligations towards them;
6. The programming of automated traffic and accidents may not be left to professional unwritten norms and/or private agreements (*volenti non fit iniuria* contracts not valid in the criminal law);
7. Based on the duty of care of users of motor vehicles (e.g. *Eagle v Chambers)*, AVs may not target pedestrians of cyclists if the option of hitting another vehicle is open, no matter if pedestrians and cyclists have behaved negligently;
8. AVs may not pick the vehicle to be hit in an emergency based on the level of safety of the vehicle;
9. The programming and monitoring of AVs behavior leading to the deliberate targeting of other vehicles and persons should be decided by a public authority, preferably a democratically elected body.

References

Alexander L, Moore MS (2015) Deontological Ethics. In: The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/spr2015/entries/ethics-deontological/

Asaro P (2011) A body to kick, but still no soul to damn: legal perspectives on robotics. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, pp 169–186

Austin JL (1956) A Plea for excuses. In: J.L. Austin (1961) Philosophical papers, edited by J.O. Urmson and G.J. Warnock. Oxford, Oxford University Press
Bonnefon JF, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. Science 352(6293):1573–1576. doi:10.1126/science.aaf2654
Christie GC (1999) The defense of necessity considered from the legal and moral points of view. Duke Law J 48(5):975–1042
Cohan J (2006) Homicide by necessity. Chapman Law Review 10(1):119–185
Dennis IH (2009) On necessity as a defence to crime: possibilities, problems and the limits of justification and excuse. Crim Law Philos 3(1):29–49
Eagle v Chambers (2004) rtr 9
Fletcher GP (2000) Rethinking criminal law. Oxford University Press, New York
Foot P (1967) The problem of abortion and the doctrine of double effect. Oxford Review 5:5–15
Gerdes JC, Thornton SM (2015) Implementable ethics for autonomous vehicles. In: Maurer M, Gerdes JC, Lenz B, Winner H (eds) Autonomes Fahren. Springer, Berlin Heidelberg, pp 87–102. http://link.springer.com/chapter/10.1007%2F978-3-662-45854-9_5
Goddard & Walker v Greenwood (2003) RTR 10
Harris J (1975) The survival lottery. Philosophy 50(191):81–87. doi:10.1017/S0031819100059118
Hart HLA (1961) The concept of law. Clarendon, Oxford
Latta v. New Orleans & N.W. Ry (1912) 59 So. 250
Lin P (2015) Why ethics matters for autonomous cars. In: Maurer M, Gerdes JC, Lenz B, Winner H (eds) Autonomes Fahren. Springer, Berlin Heidelberg, pp 69–85. http://link.springer.com/chapter/10.1007/978-3-662-45854-9_4
Matthias A (2004) The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics Inf Technol 6(3):175–183. doi:10.1007/s10676-004-3422-1
McSherry B (2002) The doctrine of necessity and medical treatment. J Law Med 10(1):10–16
Model Penal Code (1958). Tentative Draft No 8
Moore MS (1997) Placing blame: a general theory of the criminal law. Clarendon Press, Oxford
Moore MS (2010) Act and crime: the philosophy of action and its implications for criminal law. Oxford University Press, New York
Norrie A (2014) Crime, reason and history. Cambridge University Press, New York
Nyholm S, Smids J (2016) The ethics of accident-algorithms for self-driving cars: an applied trolley problem? Ethical Theory Moral Pract 19(5):1275–1289. doi:10.1007/s10677-016-9745-2
Pagallo U (2013) The Laws of Robots: Crimes, Contracts, and Torts. Springer, Dordrecht
Quinn W (1989) Actions, intentions, and consequences: the doctrine of double effect. Philos Public Aff 18(4):334–351
R v Dudley and Stephens (1884) 14 QBD 273 DC
R v Longbottom (1849) 3 Cox 439
R v Swindall and Osborne (1846) 2 Car & Kir 230
Re A (conjoined twins) (2001) 2 WLR 480
Robinson PH (1975) A theory of justification: societal harm as a prerequisite for criminal liability. 23 UCLA L. Rev 23:266–292
Ruiz v Forman (1974) 514 S.W.2d 817
SAE Information Report (J3016) (n.d.) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems
Santoni de Sio F (2016) Ethics and self-driving cars: a white paper on responsible innovation in automated driving systems. Dutch Ministry of Infrastructure and Environment Rijkswaterstaat
Singer P (1977) Utility and the survival lottery. Philosophy 52(200):218–222
Smith JC (1989) Justification and excuse in the criminal law, Sweet & Maxwell, London
Smith, BW (2015) Regulation and the Risk of Inaction In: Maurer M, Gerdes JC, Lenz B, Winner H (eds) Autonomes Fahren, Springer, Berlin pp 593–609 http://link.springer.com/chapter/10.1007%2F978-3-662-45854-9_27#page-1
Sparrow R (2007) Killer Robots. J Appl Philos 24(1):62–77. doi:10.1111/j.1468-5930.2007.00346.x
State v. Scribner (2002) 72 Conn. App 736
Taurek JM (1977) Should the numbers count? Philos Public Aff 6(4):293–316
Thomson JJ (1985) The trolley problem. The Yale Law Journal 94(6):1395–1415. doi:10.2307/796133
Thomson JJ (1986) Rights, restitution and risk: Essays in moral theory, Harvard University Press, Cambridge
Thomson JJ (1990) The realm of rights. Harvard University Press, Cambridge
United States v. Holmes (1842) 26 F. Cas. 360 (C.C.E.D. Pa. 1842)
Wood A (2011). Humanity as end in itself. In: Scheffler S (ed) On What Matters (Vol. 2). Oxford University Press, New York. http://oxfordindex.oup.com/view/10.1093/acprof:osobl/9780199572816.003.0003