

„TROLLEYOLOGY” AND AUTONOMOUS VEHICLES – MORAL AND LEGAL QUESTIONS ON THE APPLICATION OF THE DOCTRINE OF DOUBLE EFFECT¹

Dr. Lea PÓDÖR²

Abstract

The paper focuses on a classical problem of ethics and law: the doctrine of double effect (DDE). Nowadays the doctrine is more and more popular since AI-technology and super-intelligent machines have been developing rapidly. Maybe autonomous vehicles' most difficult dilemma is the following scenario: an autonomous car gets into an extreme road accident (collision) and the software should "decide" which direction-alternative to choose, but all of those possibilities end with death of human(s). This is a problem which requires a morally and legally justified answer. The principle emphasizes how to achieve a moral justification, what can be the classical cases of DDE and how to solve the most famous classical case, the trolley problem – which can be analogous to autonomous cars' collision-case. The paper also highlights whether the doctrine is relevant from the perspective of legal justification and legal solutions, too.

Keywords

autonomous vehicles, doctrine of double effect, law and morals, moral decisionmaking, moral justification of acts, intention

Introduction

This paper focuses on a classical problem of ethics and law: the doctrine of double effect (DDE). The main cause of this excursus is modern technology, namely: the autonomous vehicles. Modern technology can help everyday life in a lot of way, but there are some risks which we have to face day by day: endangerment of human life, health, body and property.³ In these dangerous situations, law has usually proved to be a very useful device to safe

¹ This essay is a part and a result of a Hungarian research project: Széchenyi István University Deák Ferenc Faculty of Law, GINOP-2.3.4-15-2016-00003.

² Assistant lecturer, Széchenyi István University Deák Ferenc Faculty of Law, Department of Legal Theory. Email: podor.lea@ga.sze.hu

³ BALOGH, Ágnes: A kettős hatás elve a büntetőjogban, in: KOVÁCS, Gusztáv – VARGA, Krisztina – VÉRTESI, Lázár (eds.): *Kettős hatás. Helyettesíthető-e az etika matematikával*. Pécsi Hittudományi Főiskola, Pécs 2015, 65–75.

humans and their property, but legal solutions are not really satisfying in every case. In connection with autonomous vehicles, moral aspects are always coming into question. We can also say that artificial intelligence (AI) belongs to a sphere which is the common part of law and morals. This is why regulation of autonomous cars requires a complex answer – but currently, it seems to be rather problematic not calming. If a question concerns the edge of law and morals, it can easily take its place in the focus of academic attention. The relevance and the actuality of analysing DDE should be explained briefly. The first question: what is the basic problem in connection with moral decisionmaking in transportation? A thought-experiment helps us to imagine the following scenario: there is an autonomous car, which gets into an extreme road accident. Just a second before the collision-moment, the software should “decide” which direction-alternative to choose, but all of them ends with serious loss, especially death of human(s).⁴ Whether there is an existing right answer-alternative to calculate death as an acceptable consequence or not? The driver’s power is assigned to the machine, namely to a software, so this entity will “decide” from case to case. Of course, technology can presume a lot of scenarios with different results and these moral dilemmas will be „decided” in advance. Can we imagine the future transportation in this way?

So, this paper’s aim is to demonstrate what is the good of DDE, how this theory could be used in moral dilemmas and the most interesting question: could it be an adequate and useful way to solve extreme moral situations caused by autonomous vehicles or not? Does the doctrine may be relevant from the perspective of law?

Viewpoint of technology

At first, it is important to explain why representatives of technology and industry believe that autonomous vehicles do not raise any moral question⁵ – in other words, this point of view is called Amoral Machine-thesis.⁶ Lawyers and technicians usually disagree when arguing about interdisciplinary relevant problems and this discrepancy is pretty much eye-catching in connection with this essay’s subject. According to technicians, the dilemma is analogous to some old and eternal problems which has not been solved yet nor by philosophers, jurists or anyone. That is the cause why there is no morally right decision to the extreme collision-problem, so the moral aspect should be simply blocked.⁷ As Héder writes: “(...) industry should (...) simply make proposals and ask for a compromise rather than chasing moral truths.”⁸

4 See the exciting moral decisions and empirical results of the famous Moral Machine-platform: <https://www.moralmachine.net/> Furthermore, an interesting survey is also available which demonstrates the worldwide results of Moral Machine; the online test has already been completed by millions of people from all over the world. See Awad – Dsouza – Kim et al., 2018.

5 Goodall asserts that the question of „right decision” is morally and legally ambiguous and there is disagreement among the experts. In an essay, he lists nine reasons why autonomous cars do not need a „machine ethics” at all. Goodall, 2014, 93–102.

6 Wendel, 2018, 133.

7 Wiseman and Grinberg, representatives of computer science, offer a trolley problem version to autonomous cars. See Wiseman – Grinberg, 2018, 105–113.

8 Héder, 2019, 46.

Transportation is based on a consensus made by society. It seems to be evident but people have forgotten about this because new types of machines have not been appeared for some time past – but a change would have called for a revision of the consensus. However, autonomous technology is something strange, something new, so sooner or later society should re-write this consensus. It means the following: reallocation of responsibility among manufacturers, designers, governments, owners of the vehicles and participants of transportation. So, the problem is based rather on calculation and not morals. All in all, industry's expectation is a much safer transportation, but first, pedestrians and other participants should accept the limitations of autonomous cars (for example eye-contact will be unnecessary to see into the "driver's" actions because intelligent vehicles do not understand this type of contact – even they do not have eyes). Algorithms are more intelligent than human drivers' brain: algorithms can calculate various scenarios and various results in advance, can make lots of plans and they also can "communicate" but in a different, newer way.⁹ Of course, the ideal ambition would be to programize autonomous vehicles like this: vehicles should be constantly able to minimize the harm and number of deaths and to save the life of passengers, pedestrians and every living creature on the roads. Would it be a solvable expectation?¹⁰

Of course, the essay does not accept this theory because we want to discover the nature of moral problems. That is why we should elaborate DDE's classical and modern readings and then we may realize how it could help (or not?) industry and legal regulation methods as well.

Traditional and modern readings of DDE. The birth of "trolleyology"

When a problem (like the one described above) proposes a choice between human lives, it can sometimes be morally solved by double effect-theory¹¹ – if the human act/decision is well tested with DDE, the act/decision may have a morally justified solution. The doctrine of double effect is quite intriguing as it supposes a flabbergasting idea: an act causing serious harm (death of – at least! – one human being) could be permissible as this harm is just the act's side effect and the agent's original aim is to achieve some good end. So the act has two effects: the good result is the intended effect and the harm is the side effect and is inevitable. There are a lot of occasions when a choice with double effect could be the only one possibility in a serious dilemma. The agent does not want to cause harm, but unfortunately she cannot avoid it.¹²

9 Héder, 2020, 154–155.

10 Nyholm– Smids, 2016, 1275.

11 The doctrine has psychological origins as well, so the human moral judgement can be described with a psychological aspect. If we are critic with this view, we can say: this aspect is an accidental byproduct of cognitive architecture. See more in Cushman's great essay: Cushman, 2014, 1–14.

12 McIntyre, 2014.

The doctrine also demonstrates how clash of moral and legal questions occurs. Originally, DDE is a principle of moral theology and was invented by Thomas Aquinas¹³ in his famous work, *Summa Theologiae*.¹⁴ Aquinas played a leading role in defining moral theology's main function and its dogmatics.¹⁵ Generally, as he emphasized, human acts have a “golden rule”: „Bonum ex integra causa, malum ex quovis defectu.”¹⁶ Besides this, the philosopher also concentrated on situations where there is no right answer as the problem always causes something wrong which is unavoidable.¹⁷ At Aquinas, the original idea of DDE refers to the following case: killing in self-defense is sometimes permitted, more precisely, killing one's assailant is justified. This act is special because it has two effects, a *prima facie* good and a *prima facie* evil effect: one which the agent intends (saving his own life) and one which is not intended but inevitable – after Aquinas, we can also say “*praeter intentionem*” (killing the attacker).¹⁸ Of course, “ (...) if a man, in self-defense, uses more than necessary violence, it will be unlawful: whereas if he repel force with moderation his defense will be lawful, because according to the jurists (...), ‘it is lawful to repel force by force, provided one does not exceed the limits of a blameless defense.’”¹⁹ According to moral theology, the application of the principle has four conditions:

1. “the act itself must be morally good or at least indifferent;
2. the good effect and not the evil effect be intended;
3. the good effect be not produced by means of the evil effect;
4. there be a proportionately grave reason for permitting the evil effect.”²⁰

These four conditions form a moral “test” and suppose that there are empirical examples which could be analysed with these conditions. In other words: casuistry means the practical aspect of theoretical dilemmas. Casuistry's²¹ subjects are moral dilemmas, more precisely cases of conscience – it seems casuistry elaborated the practical side of moral questions. These kind of problems were also analysed by Aristotle and catholic theorists like Aquinas. Casuistry's vital questions are usually dilemmas connected to life and death, such as abortion or euthanasia.²²

13 Next to Aquinas, Aristotle can be an other theorist of double effect. According to Di Nucci, the ancient philosopher wrote about various kinds of actions in *Nicomachean Ethics*. One of them, mixed action is voluntary and involuntary as well. There are external circumstances which force the agent to do something and she cannot avoid the act's bad consequences. Di Nucci, 2014a, 22–23. Aristotle brings two examples which – we can say – lookalike double effect: „the case of a tyrant” (somebody should obey the tyrant's rules because if he does not do this, his family would be put to death) and „the case of storm” (people are staying on a boat during a heavy storm; if they do not throw the goods into the water, they would die). Aristotle, 1987, 1110a9–20.

14 In book II, under question 64, Aquinas examines seven questions (objections and possible answers, too): Is it a sin to kill dumb animals or even plants? Is it lawful to kill a sinner? Is this lawful to a private individual, or to a public person or to a cleric? Is it lawful to kill oneself? Is it lawful to kill a just man? Is it lawful to kill a man in self-defense? Is accidental homicide a mortal sin? Aquinas, Thomas, Book II, q. 64.

15 Deák, 2004, 322–329.

16 Aquinas, Thomas, Book I, 18. q. 2. a. ad 3.

17 Aquinas's principle is quite interesting because catholic church emphasizes that people should pursue the good and avoid the evil. Cathcart, 2013, 67.

18 Aquinas, Thomas, Book II, q. 64.; Cavanaugh, 1997.

19 Aquinas, Thomas, Book II, q. 64. 7.

20 Mangan, 1949, 43.

21 The expression comes from the latin word „casus”, meaning „case”.

22 Szabó, 2017, 317–321.

The principle became popular in ethics and was adopted into philosophical mainstream as well.²³ What is more, it has some famous paradigmatic cases which are challenging for lawyers not just for philosophers. All of these cases show a serious problem: killing innocents. We can also say that the human act's non intended effect is other person's death. DDE has classical examples, more precisely paradigmatic cases, which are the following: killing in self-defense, abortion, euthanasia, bombing and the trolley problem.²⁴ Besides killing, the cases have one more thing in common: death could be morally (and legally) justified somehow in these extraordinary situations and death is the evil effect which is not intended at all but unavoidable. In literature, theorists elaborated the justified and non justified versions of these paradigmatic cases and have been arguing about "killing" and "letting die" for decades. I would like to demonstrate the brief descriptions and solutions of these cases in a tablet – it summarizes practicably and concisely the ethical, philosophical and legal aspects as well:²⁵

CASE	GOOD EFFECT	BAD EFFECT	JUSTIFIED VERSION	NON JUSTIFIED VERSION
Self-defense	The agent is saved	Attacker is killed	Killing in self-defense	Intended killing
Abortion	Mother is saved	Foetus is killed	Hysterectomy	Craniotomy (?)
Euthanasia	Pain relief	Death is hastened	Pain-alleviation	Intended killing
Bombing	Factory is destroyed	Innocents are killed	Strategic bombing	Terror bombing
Trolley Problem	Five individuals are saved	One person is killed	„Spur“-variant	„Fat man“-variant
Trolley Problem with Autonomous Cars	Avoiding death (or damage in property)	Death of innocent people	???	???

Development of AI calls attention again to the application of double effect, namely the new formulation of the trolley problem.²⁶ Trolley cases are known in a lot of versions and all of them have an important feature: to decide between two acts (between a good and a bad

²³ Cusham, 2014, 2.

²⁴ McIntyre, 2014.

²⁵ Di Nucci also created tablets to show the common point of DDE-cases. I used Di Nucci's tablets and completed them with my own aspects. See Di Nucci's original tablets: Di Nucci, 2014a, 29, 31.

²⁶ Nowadays it is still difficult to imagine a trolley problem-scenario with autonomous cars, because it is not a real dilemma for the majority of manual drivers. Johansson and Nilsson try to prove that the trolley-case could be interesting from the perspective of manual drivers, too. Johansson – Nilsson, 2016.

one), three conditions must be met to consider the case as a trolley case. These conditions are: “First, in trolley cases a collision is imminent and unavoidable. Second, the agent is able to choose how to distribute the harms that ensue as a result of this collision. Third, the decision situation is one of certainty. Actions carry no risk so that the agent can choose between outcomes.”²⁷

Trolley problem is usually labelled as a thought-experiment invented by philosopher Philippa Foot in 1967, in her famous essay, *The Problem of Abortion and the Doctrine of Double Effect*. Otherwise, thought-experiments are quite popular in speculative theory-making, because they are cheap and with them, scientists can prepare for future problem-solving (through answering “what if...” questions) and can elaborate various answer-alternatives.²⁸ Here are the basic thoughts of DDE and trolley problem by Foot, whose work means the birth of “trolleyology”. Foot was arguing with Herbert Hart and they discussed the problem of abortion through the double effect-theory.²⁹ Foot believed that people could not use DDE to criticize abortion.³⁰ As Aquinas, she also made a distinction between twin effects: the one is aimed, the other is foreseen but not intended; in law, they are called direct intention and oblique intention. Her original example was about a fat man stuck in a cave. “A party of potholers have imprudently allowed the fat man to lead them as they make their way out of the cave, and he gets stuck, trapping the others behind him. Obviously the right thing to do is to sit down and wait until the fat man grows thin; but philosophers have arranged that flood waters should be rising within the cave. Luckily (luckily?) the trapped party have with them a stick of dynamite with which they can blast the fat man out of the mouth of the cave. Either they use the dynamite or they drown. (...) Problem: may they use the dynamite or not?”³¹

She also writes (but briefly) about the famous trolley-scenario: imagine that “ (...) a driver of a runaway tram which he can only steer from one narrow track onto another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed. (...) why we should say, without hesitation, that the driver should steer for the less occupied track (...)?”³² This so-called basic scenario is “Spur”, as quoted in literature of philosophy. Foot has many other cases, and one of them, the „Transplant Case” is also very relevant. There are five patients who need organs and they will die if nothing happens. One day, a healthy and young man comes into the hospital; can the doctors sacrifice him as a donor to save five patients or not? The most troubling question is “why our moral reactions differ in these two kinds of cases – cases such as “Spur”, where it seems morally acceptable to take a life to save five lives, and cases such as “Transplant”, where it doesn’t.”³³ Here it is the answer: in “Spur”, somebody is redirecting an already

27 Himmelreich, 2018, 671.

28 Kovács, 2015, 85-86, 90.

29 From this view, topic of abortion is quite interesting. Foot, the English philosopher, published her essay in 1967, and in Britain, abortion was legalized by parliament in October 1967. Things were not the same in the U. S. where in 1973 a landmark case, *Roe v. Wade* made the same change as in Britain. Edmonds, 2014, 25.

30 Černý, 2020, 89.

31 Foot, 2013, 537.

32 Foot, 2013, 538.

33 Edmonds, 2014, 33.

existing threat, but in "Transplant", sacrificing the innocent man's life is a means to save five individuals.³⁴

Next to Foot, Judith Jarvis Thomson also explained the trolley case and she has more and more examples, indeed. The most famous one is the variant of the fat man case. Imagine the scenario: "George is on a footbridge over the trolley tracks (...) and can see a trolley approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. (...) the only way to stop the out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die."³⁵ And to mention one more fat man-example, here it is „Loop”: " (...) the trolley is heading toward five men who are all skinny. If the trolley were to collide into them they would die, but their combined bulk would stop the train. You could instead turn the trolley onto a loop. One fat man is tied onto the loop. His weight alone will stop the trolley, preventing it from continuing around the loop and killing the five. Should you turn the trolley down the loop?"³⁶

Philosophers say that nowadays we should imagine the classical thought-experiment not with trolleys, but with autonomous cars. The main question is constant: what is the right thing to do? The difference between intelligent vehicles and traditional vehicles is that autonomous cars do not have to cooperate with humans, the machine „decides" on its own (in a way we can say: it "decides" in advance), according to algorithms. When answering the problem, two main viewpoints are outlined: the consequentialist and the non-consequentialist approach. The first one dictates that the less fatal end is better than the more; for example, if the car kills one person instead of five, this "decision" is okay. The other approach asserts that there is a moral distinction between killing and letting die, or as Foot said: negative duties and positive duties.³⁷ As for the problem of moral responsibility, letting die is somehow better than killing because it is a passive act.³⁸ By the way, these two viewpoints are inherited in the classical literature of trolley cases.

How to test autonomous cars's "action" through DDE? What we should do is to explain the legal answer to the trolley cases: why is it allowed to sacrifice (kill) someone in these paradigmatic cases?

Some of the paradigmatic cases of double effect are known in legal theory and legal practice as well. In a way, medical activity (such as abortion, euthanasia), transportation or war could be the rare "exceptions" of life-saving obligation of the states. Especially medical activity and transportation do not mean social danger in a proper sense (of course we know, killing counts as a criminal act). Human life and right to life is a supreme value in every legal system, which means that it must not be illimitable because life is an absolute

34 Edmond, 2014, 34.

35 Thomson, 2013, 545.

36 Edmonds, 2014, 41. Critics say, „Loop case" is a very discriminating case-variant as it distinguishes innocent people according to their size (fat or skinny), so decisions could not be proper and right in this example.

37 Černý, 2020, 94.

38 Lin, 2016, 79.

value. But there are some cases when states turn a blind eye to killing – legal systems have special and divergent rules regarding these cases.³⁹ As for trolley dilemma (whether a trolley or an autonomous car we face with), it is not an exception yet but can be in the near future. Perhaps everything depends on representatives of technology (designers, manufacturers, programmers), of industry and legislators.

Trolley problem – pro and contra and its application in legal dilemmas

In literature, there are a lot of different positions in connection with autonomous cars’ trolley cases. In this section, I would like to summarize these viewpoints. Of course, theorists form two groups: lots of experts admit that trolley case is a good analogy and others think that is kind of a “dead end”. Now, what we know is not much: trolley case is not a regulated exception of killing, so it would be interesting to analyze whether DDE (as a moral aspect) could help the legal solution-searching? What can we utilise from the application of the principle?

At first, I would like to mention Di Nucci’s position, who is a well-known theorist of double effect doctrine, and who offers eight general arguments against DDE. Originally, the doctrine wanted to show the difference between intended means and merely foreseen side effects, but it failed this task. The test pays attention to an unrealistic scenario where only bad effects are taken into account and there is no real good end at all. Moreover, Di Nucci made an empirical research; he wondered how people think about the trolley dilemma? Results were surprising: most of the participants did not opt to kill the one person to save five innocents – although people usually sympathize with the utilitarian approach! There is a dilemma with the closeness problem as well because in fact, the distinction between intending and merely foreseen harm is unworkable. Fifth, Di Nucci argues with the example “Loop Variant” to prove that means are not necessarily intended. As we have seen above in the tablet, the bombing-case is a paradigmatic case and it has two versions: the terror bomber and the strategic bomber. This distinction highlights that there is no moral difference between these cases, but DDE wants to show a normative distinction with these examples. Furthermore, the doctrine is just a useless moral principle and unfortunately we will not know how to act morally right or morally permissibly. And last but not least, there is one more relevant comment from Di Nucci: people should be responsible for their non-intended acts such as their intended ones – and DDE may efface this essential obligation.⁴⁰ Regarding Di Nucci’s arguments, we can also add some general objections against the doctrine. It cannot be an appropriate solution in legal context, because it is a weak test as it is like an easy “geometrical test” and not more. Besides this counterargument, the principle is too formal and too abstract, therefore it ignores cases’ uniqueness; just works as a hypothesis and that is why it can eliminate uncertainty so easily.⁴¹ Actually, the circumstances do not exist in reality and sometimes are too extreme and unimaginable.

³⁹ To check this statement, it is important to see and read states’ regulations on abortion, euthanasia or war.

⁴⁰ Di Nucci, 2014b, 6–12.

⁴¹ Sandel, 2009, 36–39.

From ethics's viewpoint, DDE would answer the problem of verifiability but from the perspective of law, it lacks of authorisation.⁴²

We can admit: these arguments are rationale.

Hevelke and Nida-Rümelin assert that autonomous cars' extreme road accidents are differ from the paradigmatic trolley case. According to them, when autonomous vehicles come into question, we should not focus on the damage which appears in the end, "(...) when we try to determine if a decision in favour of autonomous vehicles is in the interest of one of the affected parties."⁴³

Nyholm and Smids summarize five points where the classical trolley cases and autonomous cars' trolley case differ. At first, we discuss the classical trolley problem's main features: 1) one single individual is faced by the decision; 2) the decision is an immediate one, we can say a „here and now“ decision; 3) answer-alternatives are restricted to a small number of considerations; 4) moral and legal responsibility not really matter, nor legal, nor moral responsibility is taken into account; 5) as for the modality of knowledge, the facts of the scenario are certain and unknown at the same time. If we imagine the traditional trolley-scenario with autonomous vehicles, the following factors are coming into question: 1) instead of only one person, groups of individuals face with the burden of moral decisionmaking; 2) decision-alternatives are “decided” in advance because the software is programmed with various outcomes; 3) consequently, the number of considerations is not limited; 4) both moral and legal responsibility matters; and 5) the software's „knowledge“ is characterized by risk-estimation.⁴⁴

After pro and contra aspects, it would be useful to see how does law really think about the application of DDE. When law is coming to question, it is important to understand how does it see this problem. I have already mention that in some borderline cases killing could be justified. We can add: besides the paradigmatic cases, killing is also allowed when it is committed during self-defense and necessity – these are the special causes of decriminalization. As for killing, the statement of facts is rather simple and abstract in the criminal statutes – that is because the result of the act is the relevant fact. Sometimes the need of a criterion of killing come into view – usually when something extraordinary happens and there is no appropriate or right legal (and moral) solution. I can also mention some famous cases (“hard cases”) as good examples: Regina vs Dudley and Stephens, Fuller's case of speluncean explorers, cases in connection with abortion, euthanasia or separation of siamese twins (Re A – conjoined twins case). States should not decide which are the so-called certain cases of killing, but in borderline cases states can “allow” the taking away of lives and of course there are situations where risking of life, offence of right to life can arise, too (think of abortion, euthanasia or usage of guns). Moreover, in these cases, law knows very well that agents (for example doctors) do not want to „kill“ somebody – but death is an unavoidable scenario. And what to do with autonomous vehicles? Can these be borderline cases like the other classical paradigmatic cases? I think this question should be solved in the near future such as more moral dilemmas in connection with AI technology (for example the question about their “legal entity”). But now, I would like to discuss some

42 Keenan, 2015, 16, 28.

43 Hevelke – Nida-Rümelin, 2015, 622.

44 Nyholm – Smids, 2016, 1287.

legal solutions as propositions on how to solve the question right. Probably, these answers could help a little to understand what kind of challenge legal systems have to face with.

In fact, law does not apply DDE in the traditional reading. In the XXth century, there was an intense interest from legal theorists, especially from Herbert Hart who concentrated on the principle of double effect and criticized it. Some consequences of an action are so firmly linked to the action that they are immediately and invariably accompanied by it. Here Hart touches on a problem known thanks to Foot as the closeness problem.⁴⁵ The gist: there are cases where “(...) there is a non-contingent relationship between the action and its effect. In connection with the closeness problem the question is also discussed in ethics and bioethics whether an agent can legitimately claim that he did not intend a certain consequence of an action (death), although he did intend something that is necessarily associated with it. Theorists of law frequently distinguish between direct intention and oblique intention. In Hart’s view this distinction between that which is directly intended and that which is merely foreseen as a concomitant consequence of an action is inadequately employed in the traditional doctrine of the principle of double effect.”⁴⁶

What is the problem with autonomous cars? They do not have responsibility or legal entity or cannot “decide” like a human being in the traditional reading, so now we cannot discuss their “intended and non intended decisions and actions”. So what do we have? In a figurative sense, we can imagine a trolley-scenario with intelligent cars, but not a human being will drive it or decide – instead, a machine, more precisely, algorithms will drive and “decide”. The software will understand no end of inputs and outputs, but these “questions and answers” will definitely contain new formulations of intention – it may be called the “intention-doctrine of autonomous vehicles”. Unfortunately, no one knows details of this new doctrine, so we should solve the trolley dilemma with our actual and well-known solutions (and with some new theories).

There are theorists who suggest necessity as a good direction. Personally I doubt in it; as we know it was not applicable in *Regina vs Dudley and Stephens* because of an important cause: necessity could not be a shelter of accused (who killed innocent people). And of course, necessity is always imagined as an extenuative, or as a last resort of human beings, not machines.

Philippo Santoni de Sio emphasizes that some cases are not so complex as we think. For example, we know that human life is a supreme value, consequently it would be defended better than damage in property or loss of animals. Of course, human life is incommensurable with another human life, and when matching lives comes into question, it can be solved by a consensus which is made among manufacturers and designers. Santoni de Sio concentrates on a contractarian variant where all sort of damage have a type of compensation. It seems our question is rather a contractarian problem, not of a criminal one.⁴⁷

45 Hart’s meaningful example: „If a man struck a glass violently with a hammer, knowing that the blow would break it, he would be said to have broken the glass intentionally (though not, perhaps, to have intentionally broken the glass), even if he merely wanted the noise of the hammer making contact with the glass to attract attention.” Hart, 2008, 120.

46 Černý, 2020, 86.

47 Santoni de Sio, 2017, 426–428.

Geoff Keeling elaborated a system called moral-design problem: it would be able to solve legally and morally complex situations. He criticizes Santonio de Sio, and represents a kind of utilitarian-economical approach called restricted Pareto-principle (RPP). "In collisions where 1) harm to at least one person is unavoidable and 2) a choice about how to allocate harm between different persons is required, then if there exists a unique Pareto efficient allocation of harm across different persons, then other things being equal, programming a driverless car to bring about the Pareto efficient allocation of harm is justified."⁴⁸ In a way, RPP can be a plausible reading of necessity as well, moreover, according to the theorist, RPP is a special decision-rule which can be acceptable at least three moral theories (these three are: utilitarianism, contractualism and deontological aspect).⁴⁹

Conclusion

In this short essay I tried to represent our near future's most interesting problem. In a way, the dilemma is a classical one, because "morals vs law" question is an old and eternal topic – and autonomous vehicles provide a new reading. The new trolley problem should have a satisfying solution which should be accepted both by industry and jurisprudence. Regarding the nature of the dilemma, it seems like our thought-experiment is restricted to a "mathematical example" because the question will be answered with algorithms – but of course, it is not just an easy, mechanical question, because it raises very serious moral and legal aspects and nothing is black or white, but "grey".⁵⁰ I think this problem is in connection with the idea of law. Since thousands of years, philosophers and theorists of law have been arguing about whether mathematics or argumentation could be the most appropriate one. In the past few decades, argumentation was the greater value, but in the era of AI, mathematics may occupy again its throne. What if this old-new "legal trend" will not really think of the problem's moral aspect? How will law regulate these serious problems without ethics?

References

Aristotle: *Nikomakhoszi etika*, Európa, Budapest, 1987.

BALOGH, Ágnes: A kettős hatás elve a büntetőjogban, in: KOVÁCS, Gusztáv – VARGA, Krisztina – VÉRTESI, Lázár (eds.): *Kettős hatás. Helyettesíthető-e az etika matematikával*. Pécsi Hittudományi Főiskola, Pécs 2015, 65–75.

CATHCART, Thomas: *The Trolley Problem or Would You Throw the Fat Guy Off the Bridge? A Philosophical Conundrum*. Workman Publishing, New York 2013.

⁴⁸ Keeling, 2018, 422.

⁴⁹ Keeling, 2018, 425.

⁵⁰ Orbán, 2019, 158, 163.

CAVANAUGH, Thomas A.: Aquina’s Account of Double Effect. *Philosophy*. Paper 33, 1997. Available from: <https://repository.usfca.edu/cgi/viewcontent.cgi?article=1027&context=phil>.

CUSHMAN, Fiery: The Psychological Origins of the Doctrine of Double Effect, in: *Criminal Law and Philosophy*, 10, 2014, 4, 1–14.

ČERNÝ, David: *The Principle of Double Effect. A History and Philosophical Defense*. Routledge, New York 2020.

DEÁK, Viktória Hedvig: Aquinói Szent Tamás teológiai summája és a Summa teológiája. *Vigilia*. 2004/5. 322–329. Available from: http://real.mtak.hu/24187/1/Deak_Viktoria_Hedvig_Vigilia_2004_5_u.pdf.

DI NUCCI, Ezio: Aristotle and Double Effect, in: *Journal of Ancient Philosophy*, 8, 2014a, 1, 20–48.

DI NUCCI, Ezio: Eight Arguments Against Double Effect. Proceedings of the XXIII. Kongress der Deutschen Gesellschaft für Philosophie (forthcoming). 2014b. 1–12. Available from: <https://philpapers.org/archive/DINEAA.pdf>.

EDMOND Awad, – DSOUZA, Sohan – KIM, Richard et al., *The Moral Machine Experiment*. October 2018. Available from: <https://ore.exeter.ac.uk/repository/handle/10871/39187>.

EDMONDS, David: *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us About Right and Wrong*. Princeton University Press, Princeton and Oxford 2014.

FOOT, Philippa: The Problem of Abortion and the Doctrine of Double Effect, in: SHAFER-LANDAU, Russ (ed.): *Ethical Theory: An Anthology*. John Wiley & Sons, Inc. 2013, 536–542.

GOODALL, Noah J.: Machine Ethics and Automated Vehicles, in: MEYER, Gereon – BEIKER, Sven (eds.): *Road Vehicle Automation*. Springer 2014, 93–102.

HART, Herbert L. A.: Intention and Punishment, in HART, Herbert L. A.: *Punishment and Responsibility. Essays in the Philosophy of Law*. Oxford University Press 2008, 113–135.

HÉDER, Mihály: A Contractarian Ethical Framework for Developing Autonomous Vehicles, in: ERCIM NEWS, 118, July 2019, 46–47. Available from: https://eprints.sztaki.hu/9687/1/Heder_46_30760531_ny.pdf.

HÉDER, Mihály: *Mesterséges intelligencia. Filozófiai kérdések, gyakorlati válaszok*. Gondolat, Budapest 2020.

HEVELKE, Alexander – NIDA-RÜMELIN, Julian: Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis, in: *Science and Engineering Ethics* 21, 2015, 619–630.

HIMMELREICH, Johannes: Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations, in: *Ethical Theory and Moral Practice*, 21, 2018, 8, 669–684.

JOHANSSON, Rolf – NILSSON, Jonas: Disarming the Trolley Problem – Why Self-Driving Cars Do Not Need to Choose Whom to Kill. Workshop CARS 2016 – Critical Automotive Applications: Robustness & Safety, Göteborg, Sept 2016, Sweden.

KEELING, Geoff: Legal Necessity, Pareto Efficiency & Justified Killing in Autonomous Vehicle Collisions, in: *Ethical Theory and Moral Practice*, 2018, 21, 413–427.

KEENAN, James F.: Mi a szerepe a kettős hatás elvének?, in KOVÁCS, Gusztáv – VARGA, Krisztina – VÉRTESI, Lázár (eds.): *Kettős hatás. Helyettesíthető-e az etika matematikával.* Pécsi Hittudományi Főiskola, Pécs 2015. 9–32.

KOVÁCS, Gusztáv: A puding próbája az evés. A kettős hatás elve a valóság mérlegén, in: KOVÁCS, Gusztáv – VARGA, Krisztina – VÉRTESI, Lázár (eds.): *Kettős hatás. Helyettesíthető-e az etika matematikával.* Pécsi Hittudományi Főiskola, Pécs 2015, 84–90.

LIN, Patrick: Why Ethics Matters for Autonomous Cars?, in: MAURER, Markus – GERDES, Christian J. – LENZ, Barbara – WINNER, Hermann (eds): *Autonomous Driving. Technical, Legal and Social Aspects.* Springer 2016, 69–85.

MANGAN, Joseph T.: An Historical Analysis of the Principle of Double Effect, in: *Theological Studies*, 10, 1949, 1, 41–61.

MCINTYRE, Alison: The Doctrine of Double Effect, in: ZALTA, Edward N. (ed.): *The Stanford Encyclopedia of Philosophy*, 2014. www.plato.stanford.edu/entries/double-effect.

NYHOLM, Sven – SMIDS, Jilles: The Ethics of Accident-Algorithms for Self-Driving Cars: and Applied Trolley Problem?, in: *Ethical Theory and Moral Practice*, 19, 2016, 5, 1275–1289.

ORBÁN, Endre: Amikor a nézőt rendezik: morális dilemmák a Terror árnyékában, in: *Fundamentum*, 2019, 3–4, 155–167.

SANDEL, Michael J.: *Mi igazságos... és mi nem? A helyes cselekvés elmélete és gyakorlata.* Corvina, Budapest 2009.

SANTONI DE SIO, Philippo: Killing by Autonomous Vehicles and the Legal Doctrine of Necessity, in: *Ethical Theory and Moral Practice*, 2017, 20, 411–429.

82	ARTICLES Lea PÓDÖR	„Trolleyology” and autonomous vehicles – moral and legal questions on the application of the doctrine of double effect	
----	-----------------------	---	--

SZABÓ, Miklós: A jogeset dialektikájához. A kazuisztika, in: TAKÁCS Péter (ed.): *Unitas Multiplex. Ünnepi tanulmányok Szigeti Péter 65. születésnapjára*. Gondolat Kiadó – SZE DFK, Budapest-Győr 2017, 316–327.

THOMAS Aquinas,: *Summa Theologiae* I-II. Available from: <https://www.newadvent.org/summa/>.

THOMSON, Judith Jarvis: Killing, Letting Die, and the Trolley Problem, in: SHAFER-LANDAU, Russ (ed.): *Ethical Theory: An Anthology*. John Wiley & Sons, Inc. 2013, 543–551.

WENDEL, W. Bradley: Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles, in: *California Western Law Journal*, 55, 2018, 1, 129–163.

WISEMAN, Yair – GRINBERG, Ilan: The Trolley Problem Version of Autonomous Vehicles, in: *The Open Transportation Journal*, 12, 2018. 105–113.

Author

Dr. Lea Pődör
Széchenyi István University, Győr,
Deák Ferenc Faculty of Law and Political Sciences
Department of Legal Theory
9026 Győr, Áldozat u. 12.
Hungary
podor.lea@ga.sze.hu