

Prof. Dr. hab. Marek Piechowiak
Department of Theory, Philosophy
and History of Law
SWPS University of Social Sciences and Humanities
(SWPS Uniwersytet Humanistycznospołeczny)
mpiechowiak@swps.edu.pl

Review of the Doctoral Dissertation

Michael P. Musielewicz, *On the Application of Norms within Driverless Cars*
(January 26, 2019, pp. 174)

According to Abstract the main issue which is considered in the dissertation is “the question of whether driverless cars can be bearers of norms and are capable of being normative agents who can follow both legal and ethical norms” (p. ii). The Author gives positive answer to this question and, more specifically, aims at establishing “what sort of ethics applies to driverless cars” (p. 3) or – rather – what sort of ethics applies best to driverless cars. Answering this question Musielewicz proposes to apply a target centred virtue ethics.

The Author starts with the explanation what autonomous cars are and a survey of the current regulatory legal framework adopted in different countries. Then he argues that driverless cars are normative agents, both in legal and moral sense (they are addressees of both legal and moral norms) (p. 35). The next step leads to the scrutiny of different ethics which could possibly be applied to these cars. Finally, considering difficulties which the various accounts of ethics encounter the Author proposes his solution, advocating in favour of a target centred virtue ethics.

Consequently, the dissertation consists of four main chapters – (2) Autonomous Cars in the World, (3) Normative Agency for Artificial Agents, (4) Ethics and Artificial Normative Agents, (5) Difficulties in Standard Account and Solution, and also (1) Introduction, (6) Conclusion and Bibliography.

The overall structure is logical. The internal structure of each editorial unit is also well-ordered: the reader is informed what has already been established and what issues will be analysed afterwards. The Author expresses his thoughts in a clear way. Similarly, he also discusses others' views in a very comprehensible manner. The distinction between the Author's own views and those which are only referred to is very precise. His selection of relevant standpoints is very accurate and contributes to

the smoothness of presented arguments. In view of the richness of the relevant academic works in this field such a selection was a task requiring a deep understanding of the issues at stake.

Musielewicz took into consideration not only scientific but also popular literature. I am very much in favour of such an approach. It is quite rare and certainly broadening of the perspective helps in building a more comprehensive inquiry.

The bibliography contains 93 entries – certainly not an astounding number, nevertheless sufficient for developing the argumentation. I am certainly not an enthusiast of the method – adopted by the Author – of ordering bibliographical entries following the sequence in which they were referred to and not alphabetically. It makes checking if a given work has been taken into account difficult. In general, however, the thesis is correct from the formal point of view.

In his main line of argumentation, the Author examines, in a very interesting way, differences between the notion of virtue applied to autonomous cars and the notion of virtue developed in the “traditional” ethics, especially by Aristotle. In opposition to the views widely shared in scholarship on virtue ethics applied to artificial agents, he rightly points out the lack of possibility of reflecting upon overall end of artificial agents. In his argumentation he aptly refers to Aristotle showing deficiencies of the approach developed by Klaus Diepol and Nicolas Berberich (“The virtuous machine – old ethics for new technology?”). Musielewicz concludes that the driverless cars “only function and act within the sphere within which they have been trained” in accomplishing specific targets. This leads to the principal claim of his thesis, that the “target-centered virtue ethics” is suitable for considering normative framework for programming autonomous cars. The main argumentation for this claim is well established.

It has to be observed that in the course of his thesis Musielewicz sometimes promises his readers too much. He makes a clear distinction between – as he calls it – two horns which are considered in his analyses of the “behaviour” of driverless cars. One is based – as he writes – on legal norms, the other on non-legal or moral norms (“non-legal or moral behavior”, 35). He makes a reservation that both domains are interrelated, nevertheless the distinction itself might be misleading. In Chapter 2 he provides a survey of the contemporary legal regulations concerning usage of autonomous cars as a particular kind of vehicles which operate on the public roads. These regulations do not repeat many legal norms which prescribe or prohibit certain behaviour in the public domain nor norms concerning ordinary cars which apply also to the autonomous cars. His survey of legal regulations in different countries actually does not contribute much to his main argument which aims at examination of virtue ethics as a general normative framework applicable to autonomous cars. This survey informs mostly that there are legal regulations explicitly related to the usage of autonomous cars in a given

country, but the Author provides only a basic information about the content of these regulations (sometimes there is no information about the content at all). Actually, the main conclusions from the survey are that the legislators do not ascribe legal personhood to autonomous cars and that the legal responsibility for the behaviour of these cars rests on their users or producers. The considerations in Chapter 2 devoted to the legal regulations are probably the weakest part of the thesis. It is also true that the importance of these considerations for the main argument is quite limited.

Nevertheless, it remains unclear what is the relation between laws analysed in Chapter 2 and the laws analysed by the Author in Chapter 3 where – according to the Author's declaration – the legal "horn" of normative agency is at stake. In this part such legal offenses as e.g. manslaughter or criminal vehicular homicide are taken into account. It still remains unclear what is the essential difference – from the point of view of programming autonomous cars – between norms identified in law and moral norms. On page 37 the Author suggests that the difference rests on the origin of norms – laws originate from the state, morals from moral authority. He introduces there also the third kind of norms – social norms which originate "from the customs and mores of a given society". Though he hardly takes advantage of these differentiations.

The distinction between legal and moral sphere is easily comprehensible if supererogatory obligations (actions) are at stake but this issue is not a central one in the thesis. The problem with distinguishing the two "horns" becomes apparent when the Author considers the virtue of justice. He explores Aristotelian notions of legal justice and distributive justice. The Aristotelian legal justice is characterised by him as a universal justice – and it consists in acting lawfully (Musielewicz mistakenly quotes in this context Aristotle's explanations related to particular justice). He overlooks that for Aristotle the truly universal justice is equal to moral excellency as such. He also fails to notice that legal justice as based on general norms sometimes requires correction by *epikeia* which allows to take into account the specificity of a particular situation in which conflicting goods are at stake. *Epikeia*, as Aristotle writes, is better than legal justice but it is not better than the (universal) justice, it is a correction of legal justice performed to obtain the justice itself. Coming back to the distinction between law and morality, it seems that the Author considers problems to be moral, when they arise from the situations in which elementary goods are in conflict and in which strictly applying legal norms, which from their nature are general, is problematic. If it is the case the criterion of distinguishing that what is lawful from what is morally correct seems to be based on misunderstanding what law is about, and Aristotle can be helpful in clarifying this issue. Although compliance with legal norms which are general is called by him "legal justice" Aristotle is far from taking the position that such understanding of justice is sufficient in law. *Epikeia* is an important part of applying law. Finding a proper (right, just) solution when the collision of fundamental goods

(norms, interests) is at stake is essentially a legal matter. The so-called hard cases are legal cases, although moral issues are then sometimes (but not always) also at stake.

The Author's deliberations about Aristotelian approach to justice certainly need further refinement. When talking about particular justice he pays some attention to distributive justice but he does not mention the commutative one. The particular justice, as it is understood by Aristotle, is in opposition to gaining profits at the expense of others. This issue is relevant e.g. when programming "smart" (in pejorative sense) cars is at stake, cars which would be more effective but at the expense of lower effectiveness of other cars.

Sometimes readers may feel disappointed because of the extent to which the author applies traditional concepts and ethical distinctions to analyses regarding artificial agents such as autonomous cars. For example, he convincingly argues that autonomous cars meet "the well-established criteria for being an agent within the computer science community" (37), though he does not examine the critiques launched outside this community. Specifically, he does not consider to what extent free will is constitutive for being an agent and a normative agent in particular. It may be one of the reasons for blurring the boundary between being a moral subject and being a subject of law. For being a moral subject free will and also a possibility to reflect upon an overall end of their life seems to be necessary. This is not the case when subjects of law (in general) are considered, which - to some extent - can be regarded as normative agents, and they do not have to be natural persons. The question of will is present when the two main conceptions of rights are considered – will theory and interest theory. This is, however, a different issue, related mostly to how aims of actions (good) are constituted and not to deciding by the agent whether these aims (constituted this or another way) are to be realised by his action.

The Author's reasoning comprises also critical considerations of ethics of deontological and – mostly – of utilitarian (consequentialist) type. It proves that he is highly skilled in both in analytical and synthetic inquires. Discarding deontological approaches, based on very short analysis, seems to be a little bit hasty. He mentions that the attempts to ground the managing of the behaviour of driverless cars in deontological approaches are actually not present in the scientific literature. Looking for solutions in the domain of virtue ethics is, then, quite natural and consistent with the main line of argumentation in the thesis. However, considering the possibilities provided by deontological ethics, in spite of the dominant tendencies, deserves more attention. Since self-learning processes are immanent in programming autonomous cars and these processes aim at utmost possible effectiveness in accomplishing certain tasks, it is quite possible that the machines will find such means to achieve these tasks which are not acceptable from the moral point of view. For example, it

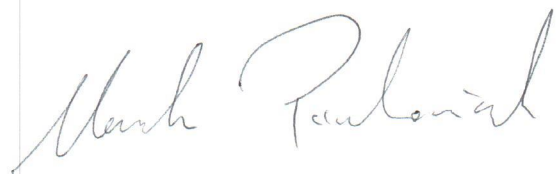
seems to be morally unacceptable to produce autonomous cars (and to offer them at a higher price) which are more efficient because of their ability to cheat other cars they are interacting with or to use algorithms enabling them to get faster to their destination but at the expense of slowing down the traffic in general. Therefore, an approach which combines virtue ethics with elements of deontological ethics seems to be a promising one, or at least worth considering. It has to be borne in mind that direct responses to some shortcomings of utilitarian ethics are provided by deontological rather than by virtue ethics.

Nevertheless, while considering these questions I, personally, would hesitate to talk about morality of autonomous cars directly and would restrict the use of the term "morality" or "moral" to the programming of these cars or to using them by humans who are free moral agents and are able to reflect upon an overall aim of their actions and of them themselves, and not only upon particular tasks which could be accomplished by cars.

Reflecting on the shortcomings of the thesis, the lack of any consideration of dignity as the fundamental moral and legal concept is striking. The legal dimension of this concept is much less visible in the common law tradition than in the continental legal culture, and the Author is certainly better acquainted with the former. Nevertheless, dignity is in the centre of many ethical traditions. Paying attention to it could provide an important insight into the normative relations between autonomous cars and their users. I must admit, however, that the problems which are usually posed when dignity is considered, are also present in Musielewicz's thesis in connection with the trolley problem which – rightly – is one of central problems which he exploits while criticising utilitarian approaches.

Though the dissertation bears imperfections, it nevertheless, as a whole, certainly meets the requirements set for doctoral dissertations.

Poznań, 19-6-2019

A handwritten signature in cursive script, reading "Mark Fendrich". The signature is written in dark ink and is positioned to the right of the date.