Do Automated Vehicles Face Moral Dilemmas? A Plea for a Political Approach

Javier Rodríguez-Alcázar

Lilian Bermejo-Luque

Alberto Molina-Pérez

ABSTRACT

How should automated vehicles (AVs) react in emergency circumstances? Most research projects and scientific literature deal with this question from a moral perspective. In particular, it is customary to treat emergencies involving AVs as instances of moral dilemmas and to use the *trolley problem* as a framework to address such alleged dilemmas. Some critics have pointed out some shortcomings of this strategy and have urged to focus on mundane traffic situations instead of trolley cases involving AVs. Besides, these authors rightly point out the political nature of the most interesting debates involving AVs. However, in our view, they do not offer an adequate account of the distinction between ethics and politics and still see their proposals as contributions to the *ethics* of AVs.

We argue that many of the interesting questions about how AVs should behave, both in emergency and other situations, are of political, and not moral, nature. This view is based on a conception of politics and political normativity that we have developed elsewhere and that we call "political minimalism". Additionally, we show that this proposed perspective has significant consequences for the design, management, and regulation of transport systems.

Keywords: Automated Vehicles; Trolley Problem; Ethics; Politics; Political Minimalism; Political Realism; Political Moralism.

1. Human drivers and AVs

Imagine a driver who, after a sudden brake failure, had to choose whether to run over a group of pedestrians or crash the car against a wall. In the latter option, people inside the vehicle would be put at risk. The choice would be difficult even if the driver had time to meditate carefully on her or his decision. In real-world circumstances, drivers do what they can. In most cases, they experience a mixture of hesitation and dissatisfaction concerning the moral goodness of what they did.

Now imagine a similar scenario[1] involving a driverless automated vehicle (AV).[2] The similarity between the two settings led some authors to talk about the moral dilemmas faced by AVs and wonder what the morally correct responses to such dilemmas would be. For instance, there is an article in *Scientific American* with the title "Driverless Cars Will Face Moral Dilemmas" (Greenemeier 2016), while Bonnefon, Shariff, and Rahwan (2016) write: "some crashes will require AVs to make difficult ethical decisions in cases that involve unavoidable harm".[3] On what grounds are AVs supposed to make such decisions? The attempt to answer this question has prompted increasing literature in the last few years (see, for instance, Moor 2006; Goodall 2014 and 2016; Deng 2015; Gerdes and Thornton 2015; Kirkpatrick 2015; Purves, Jenkins and Strawser 2015; Bonnefon, Shariff, and Rahwan 2016; Etzioni and Etzioni 2016; Greene 2016; Lin 2016; McBride 2016; Awad, Dsouza, Kim, Schulz, Henrich, Shariff, et al. 2018).

Authors dealing with the question often resorted to the *trolley problem* family of scenarios (Lin 2013; Nyholm and Smids 2016; Gogoll and Müller 2017). These are thought experiments used to check moral theories against our moral intuitions, and it is tempting to resort to them to address certain moral dilemmas —perhaps even more tempting when such dilemmas involve objects in motion. In the classical version, a driver has to choose whether to leave a runaway trolley that is barreling down the railway tracks to continue its way and kill five anonymous people, or to divert the trolley onto a sidetrack instead, causing the death of one equally anonymous person (Foot 2002). By replacing the trolley and the driver with an AV governed by software, researchers can design varying scenarios and wonder what

---

1 It might be argued that the use of AVs will completely prevent the occurrence of crashes, but this is an unlikely prospect. Goodall (2014) provides a list of reasons why accidents may persist, even if fully automated vehicles do not have to share the roads with human drivers.

2 For reasons that will become apparent later on, we prefer the expression "automated vehicles" to the more common "autonomous vehicles". By "automated vehicle" (AV) we understand a vehicle that requires no human supervision or intervention for operating. More precisely, the term refers to Level 5 vehicles in the autonomous vehicle classification scheme provided by the Society for Automotive Engineers (SAE 2018).

3 Yet another example: The National Science Foundation (USA) has awarded a grant for the project "Ethical Algorithms in Autonomous Vehicles", which searches "constructive answers to questions about autonomous vehicles, translate them into decision-making algorithms for the vehicles and then test the public health effects of those algorithms under different risk scenarios using computer modeling" (Weinberg 2017).

the morally correct solution would be in each case. Emergencies like these are treated as moral dilemmas in which the AV has to choose the lesser of two evils.

One variant of this strategy to determine how AVs should behave in different circumstances is to survey the public's answers to each dilemma. This is the essence of MIT's *Moral Machine* project, "a platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars". [4] Volunteers taking part in the experiment are confronted with a series of emergencies and asked what the AV ought to do. For instance, subjects are asked whether an AV with sudden brake failure should continue ahead, killing two men, one little girl and one boy that are flouting the law by crossing on the red signal, or rather swerve and kill three elderly people and one man that are abiding by the law by crossing on the green signal. By 2018, the *Moral Machine* platform gathered 40 million decisions from people in 233 countries and territories (Awad, Dsouza, Kim, Schulz, Henrich, Shariff, et al. 2018).

Keeling (2020, p. 294) characterizes the *trolley cases* in the context of AVs as those cases in which: "(1) the AV must choose one of two actions; (2) the AV knows what the consequences of each action will be; (3) each action imposes a distribution of benefits and burdens over at least two affected parties; and (4) the interests of these parties are jointly unsatisfiable." Several authors have rejected that the trolley cases are relevant to the selection of the values that ought to be reflected in the AVs' decision-making algorithms. Their rejection is based on different arguments, among which we shall focus only on the following: trolley cases intend to provide a moral answer to what is ultimately a political problem.[5]

Himmelreich offers a version of the latter argument. He writes: "trolley cases are taken to be an issue of morality. But we think that this locates the problem on the wrong level. Instead, solutions are called for on the level of politics. Whereas moral philosophy is a reflection on individual conduct, political philosophy is a reflection on social arrangements before the backdrop of substantive disagreement" (Himmelreich 2018, p. 676). Himmelreich concludes that, although trolley cases may be useful in certain contexts, their usefulness is very limited when it comes to the ethics of AVs. His positive stance has two components. First, a focus on mundane traffic situations ("such as approaching a crosswalk with limited visibility, making a left turn with oncoming traffic, and navigating through busy intersections", Himmelreich 2018, p. 678) instead of artificial, unlikely trolley cases. Second, an approach to the ethical problems posed by AVs as a matter of social choice that

4 See http://moralmachine.mit.edu/. The full title of the project, according to its website, is: "Moral Machine-Human Perspectives on Machine Ethics".

5 For discussions of other objections to the strategies based on the trolley cases, see Nyholm and Smids (2016) and Keeling (2020). In a talk at the University of Granada ("The Use and Abuse of the Trolley Problem: self-driving cars, medical treatments, and the distribution of harm", 5 June 2019), Frances Kamm provided some reasons why cases of AVs crashes such as the ones mentioned above should not be treated as instances of the trolley problem.

would require a political solution rather than a matter of individual choice that would require a moral solution (Himmelreich 2018, p. 681).

Roff (2018) agrees with Himmelreich for similar reasons on trolley cases' uselessness for developing the ethics of AVs. She, too, believes that mundane cases are more interesting, because AVs normally "will be making sequential decisions in a dynamic environment under conditions of uncertainty", instead of a single decision —taken with knowledge of the consequences of each alternative course of action— that is characteristic of the trolley cases. Roff also coincides with Himmelreich in the need of placing the issue under a political light. She points out that AVs need to learn to make trade-offs in real situations between social values, such as "privacy, security, trust, civil and political rights, emotional well-being, environmental sustainability, beauty, social capital, fairness, and democratic value", whose relative weight would have to be assessed politically.

Remarkably, though, despite their frequent reference to social choices and social values, both Himmelreich and Roff present their respective proposals as contributions for a better *ethics* of AVs. Himmelreich (2018, p. 670) summarizes his positive outlook as follows: "We argue for the ethical relevance of mundane traffic situations". This hesitation between the political and the moral perspective also appears when he approaches practical cases. For instance, the question of whether an individual who is late to a meeting would be allowed to pay to override some safety features of the AV is treated by Himmelreich (2018, p. 680) as an "ethical issue". As it will become apparent later on, we think that while the question of whether an individual should pay in such circumstances is a moral question, the question of whether individuals should be allowed to alter the algorithms of their AVs is not a moral question but a political one. Even clearer is our judgment on a second example provided by Himmelreich (2018, p. 681), concerning the value trade-offs faced by human drivers in mundane traffic situations (Himmelreich mentions safety, mobility, efficiency and environmental impact among these values). We claim that, while the decision taken by one individual driver on how to prioritize such values constitutes a moral issue, the question concerning the trade-offs among the same values in the legislation regulating the software for AVs is strictly political.

These authors' hesitations suggest that a clear understanding of the distinction between ethics and politics is missing in their otherwise insightful proposals. In what follows, we will argue that, although there is room for an ethics of AVs (and even for the discussion of trolley cases within it) which is related not only to the individual decisions of all the relevant actors (software engineers, consumers, lawmakers, carmakers, and others) but also to the elucidation of people's moral intuitions to make them compatible with AVs behavior, the question of the values that ought to guide the design of AVs algorithms and the question of how to adjudicate the unavoidable trade-offs among them are political questions that are better addressed using political instead of moral criteria. We purport to provide such political criteria by presenting a particular conception of political normativity and of the relationship between ethics and politics that underlies it.

2. Ethics and politics

2.1. Political moralism and its shortcomings

Kant (1781/1998, p. 677) famously stated: "All interest of my reason (the speculative as well as the practical) is united in the following three questions: 1. What can I know? 2. What should I do? 3. What may I hope?" There is no need to stress the individualistic bias in the formulation of all three questions: the three problems that all philosophy ultimately aims to solve are formulated as three questions that a sole, rational (and male, incidentally) thinker addresses to himself. Regarding the second question ("What should I do?"), Kant (1781/1998) describes it as "merely practical" and "moral". Thus, an outcome of Kant's enlightened individualism is the reduction of practical reason to morality. Any collective subject (any *community*) has vanished as a questioning subject.[6] As Williams (2005) points out, for Kant, as for most thinkers in his philosophical tradition, political philosophy becomes a chapter of *applied ethics*. This is the core thesis of what Williams calls "political moralism", a philosophical tradition also discussed by Larmore (2013).

Most contributions to the ethics of AVs take for granted political moralism. All the treatments of the topic resorting to trolley cases are moralistic indeed, but even Himmelreich (2018) and Roff (2018), who advocate "political" approaches, seem to incur in a reduction or subordination of politics to ethics when they continue to see their proposals as contributions to the ethics of AVs.

What is wrong with political moralism? One major difficulty is that political moralism ultimately assigns to ethics the solution of political debates. Political realists objected that ethics might not be of much help, given the ubiquitous disagreement among people on their conceptions of the right and the good (Larmore 2013, pp. 277–280; Rossi and Sleat 2014, p. 691). Since these conceptions include moral ones, morality would offer rather a puzzle than a solution to political dilemmas, namely, the puzzle of how to handle the plurality of moral convictions that coexist in many societies (Galston 2010, p. 400). Defenders of political moralism replied that it is doubtful that political principles are better suited than moral principles to resolving disagreements if by "resolving" we understand not the *de facto* sense but a normative sense, i.e., the sense of yielding an answer as to which party is right (Leader Maynard and Worsnip 2018, pp. 768–772). To this argument, Jubb (2019, p. 364) has retorted that, although a moral evaluation of political processes and outcomes is always possible, this does not entail that such evaluation is the most relevant. Particularly, given that political puzzles often develop in the absence of high levels of moral agreement, a straightforward moral solution to such problems would not be of much help.

---

6 On the individualistic character of Kant's approach and how liberalism inherits it, see Newey (2010, p. 456).

2.2. Political minimalism

Our stance is that the question of whether ethical or political problems are the most difficult to solve is not relevant to distinguishing between morality and politics. The fundamental point is that morality and politics respond to different questions (Author1 2017a and 2017b; Author2 and Author1 forthcoming). More precisely, we claim that the realm of morality arises when individuals confront themselves with the question "What shall I do as regards other people's interests as ends of their own and not mere means to my interests?" The line that goes from the morally supererogatory to the morally evil is then the line that goes from the maximal prioritization of other's interests over my own to the maximal prioritization of my interests over other people's. In our view, politics is a constitutively normative practice and its constitutive aim is answering the question "What shall *we* do as regards our ends and the means available to achieve them?" Thus, political normativity would be a matter of collective instrumental rationality in that good political decisions are good responses to that question. Accordingly, people are involved in either political or moral deliberation as long as they take care of answering one or another question.

Political moralists are right in vindicating the need to resolve political problems in the normative sense. However, this does not entail that the resolution of a political debate has to come from morality, as far as there are grounds for accepting the existence of a distinctively political kind of normativity.[7] It is noteworthy that our distinction between moral and political perspectives does not require a clear-cut distinction between moral and political values, as is usual in the debates between political realists and political moralists. Williams (2005) situates the debate in these terms by stating that there is a non-moral condition for politics to be possible, and this condition is order and stability. This would be the criterion to tell good from bad politics. Political moralists reply that if political order is to be distinguished from sheer domination, as Williams wants, this demand has to be qualified using moral values, like freedom and equality. This is what Erman and Möller (2013 and 2015) call the "ethics first premise". Realists like Jubb and Rossi (2015) deny that freedom and equality are always moral values and contend that if realists invoke moral values, these would be filtered through the political goal of providing order and security. We believe that focusing on whether or not values such as freedom and equality are moral values is misguided. These two values, like many others, can be seen as either moral or political depending on the question that is answered when they are invoked as reasons for action.

The main claim of this article is that, although individual drivers may face moral questions in trolley cases, or other contexts involving an individual decision ("what

7 For an account of the sources of political normativity, which is beyond the scope of this paper, see Author2 and Author1 (forthcoming). An overview of our general strategy to deal with the normativity of constitutively normative practices can be found in Author2 (2011, ch. 2); also, in Author2 and N. Villanueva (forthcoming).

shall I do...?"), the representation of AVs as moral agents leads us astray from good decision-making.

2.3. Political minimalism and political realism

The distinctively political approach to the question of AVs' behavior places our stance in the vicinity of political realism. Political realists typically contend that politics is autonomous from morality and they resist political moralists' tendency to see political philosophy as a chapter of applied ethics (Williams 2005). Besides, most political realists would agree with us that there is an autonomous political normativity whose force is not derived from moral normativity.[8] As far as the regulation of AVs is concerned, it might be possible to arrive at similar conclusions as those we defend in this paper by adopting a realist framework. Nevertheless, to defend our approach from some of the objections affecting political realism, we will explain briefly the main differences between political realism and our background meta-political theory; i.e., what we call *political minimalism*.

In the first place, realists (like moralists) tend to provide "stronger" characterizations of the political than ours, which results in an unjustified restriction of the scope of politics. For instance, Sleat (2016, p. 33) claims that disagreement and conflict are constitutive of politics, and adds that the permanent presence of such disagreement entails that the distinction between one group (the rulers) and others (the ruled), linked by legitimate coercion, is also a necessary component of politics. Yet, in our view, it is conceptually possible to conceive of instances of political practice that are cooperative instead of conflictual, and of political communities where the distinction between the rulers and ruled is not relevant (Author2 and Author1, forthcoming). Counterexamples to the realist's usual characterization of the political can be found both at the empirical level (experiences of pre-state societies and of what we treat as small political communities, like a university department) and among normative political proposals (e.g., some anarchist projects). The presence of both conflict and hierarchy in political communities is a contingent empirical fact —recognizable, indeed, in many societies—, but conflict and hierarchy do not qualify as constitutive elements of an acceptable definition of the political.

A second and deeper disagreement with political realists lies in our divergent ways of thinking of the *source* of political normativity. On this issue, different variants of political realism provide different answers. Rossi (2019) classifies such variants into three kinds: ordorealism, radical realism, and contextual realism.

Ordorealism would be exemplified by Williams (2005). Foreseeing the objection that political realism might be normatively powerless and might foster mere descriptions of the *status quo*, Williams puts forward a normative criterion: the "Basic Legitimation Demand" mentioned earlier. For Williams (2005, p. 4), this

8 It is doubtful, though, that this is the case with some political realists who adopt a more descriptive, not normative, approach to political theory. See, for instance, Geuss (2008).

Demand is met when a state can provide an acceptable solution to the 'first political question,' i.e., the securing of order. In Williams' words, this question would be "'first' because solving it is the condition of solving, indeed posing, any others" (2005, p. 3). Against this purported solution to the problem of the source of political normativity we have argued (Author1 2017a, pp. 732–735) that, with equal conviction, utilitarians might claim that not dying of starvation is a necessary condition for the possibility of political life, while Kantians might argue that freedom is a necessary condition for seeing a group of living creatures as a political community, and not as a mere herd of animals. So, both traditions could defend, with arguments similar to Williams', their respective, moralistic sources of political normativity. Our positive proposal is minimalistic in the sense that, instead of espousing any instrumental goal (either moral or non-moral) as the ultimate criterion for political evaluation, we adopt as criterion the ultimate ends pursued by each political community. Such ends may roughly coincide with most standard lists of basic human needs, and with the list of goals jointly postulated by both political realists and political moralists down the centuries (order, wellbeing, freedom, justice...), but instead of selecting one of them as *the* legitimating political goal, political minimalism allows each political community to select and rank its own set of ultimate ends, and describes good politics in terms of success at promoting such ends.

In turn, "radical realism" would find the source of political normativity in a form of ideology critique (Rossi 2019). The role of the political philosopher would be to criticize, on epistemological grounds, the legitimation stories advanced by others. Rossi (2019, p. 646) uses an illuminating analogy between ideology critique and Michelangelo Buonarroti's famous description of his work as a sculptor: sculpting is to remove marble until the figure inside emerges. But, as Rossi himself acknowledges, in ideology critique there is no preparatory study or sketch of the outcome. However, perhaps the most important element in the complete quote of Buonarrotti's phrase (for anyone not wishing to turn the phrase into a joke) is Michelangelo's reference to the identity between the concept in the artist's mind and the concept hidden in the marble, a concept that only "a hand that obeys the intellect can discover" (quoted by Rossi 2019, p. 638). Ideology critique is a necessary work indeed, analogous to the work of the stonecutter that removes the defective portions from the marble block before sending it to the sculptor. Yet ideology critique does not provide everything that one would expect from political philosophy, much in the same way in which stonecutting is not equivalent to sculpture. In a full-fledged, normative political philosophy we need a normative model, as the sculptor needs a project.[9]

9 Otherwise, the challenge for a radical realist would be better described not by Michelangelo's phrase but by another sculpture story that is popular in Spanish-speaking countries, guarding against the dangers of working without a plan. One day, someone asked a busy sculptor what piece was he about to produce. The anonymous artist, who was specialized in religious sculpture, replied that he did not know: "if the piece would happen to be bearded, it would be a Saint Anthony the Abbot; otherwise, the Virgin of the Immaculate Conception" ("si sale con barba, San Antón, y si no, la Purísima Concepción").

Finally, it might seem that the context-dependence element in our stance (i.e., the reference to the changing set of ultimate goals endorsed by each political community) places us closer to what Rossi (2019) calls "contextual realism". Contextual realists claim that the legitimation of political principles depends on the interpretation of the aims of particular political institutions and practices that exist as a matter of fact in a society (Rossi 2019, p. 642). Sangiovanni (2008) distinguishes two kinds of practice-dependence in the justification of first principles of justice: while for *cultural conventionalists* like Walzer (1983), culturally contingent values and meanings would determine first principles of justice in each society, *institutionalists* like Sangiovanni himself hold that the nature of shared political institutions, such as the modern state, provides the reasons for endorsing certain principles of justice (Sangiovanni 2008, p. 138). Rossi (2012) defends a similar stance but focusing on *legitimacy* instead of justice. Yet, a major challenge for both kinds of contextual realism is to explain how contingent values and institutions, which are supposed to be the source of political justification, are politically justified in turn. As regards Argumentation Theory, Author2 (2011, pp. 41–44) has described this challenge as a particular case of Agrippa's well-known trilemma: the attempt to justify a normative model seems to be doomed either to a vicious circle (if, in the attempt to justify the model, one uses the same criteria postulated by the model), an infinite regress (if the criteria are justified by appealing to other criteria that need further justification, and so on), or to an arbitrary stop in the regress. In our view, not only contextual realism but also Cohen's (1989 and 2008) moralistic attempt to ground justice on practice-independent intuition fall in the latter kind of problematic strategy.

2.4. Politics as a constitutively normative practice

In a nutshell, our way out the trilemma translates into Political Philosophy the solution developed by Author2 (2011, ch. 2) for normative models of Argumentation Theory: the idea is to think of these models as descriptive of a practice that is constitutively normative. Thus, for example, if whatever counts as argumentation counts as an attempt at showing a target-claim to be correct, then good argumentation will be argumentation showing a target-claim to be correct. Similarly, political philosophy would be a normative endeavor inasmuch its main goal is not to describe a particular, culturally dependent, instance of political practice, or a certain kind of political institution, but to describe a practice, politics, whose constitutive aim is to provide good responses to the question "What shall *we* do as regards our ends and the means available to achieve them?" If our description of politics as such constitutively normative practice is correct, then good politics will be politics that adequately respond to such question. If it is understood this way, politics would be an exercise of collective instrumental rationality, to be distinguished from both morality and individual instrumental (prudential) rationality.[10]

10 We do not think that the same strategy to explain the source of political normativity could be used to explain the source of moral normativity, because we do not think that morality is a constitutively normative practice. See Author2 and Author1 (forthcoming).

Two last comments are necessary at this point. First, it is important to distinguish between the normativity of politics and the normativity of political philosophy. The deliberation on what are the ultimate ends of a community is part of politics but, we claim, it is not part of political philosophy. Philosophy just tells us that norms, or institutions, are justified if they are good means for those ends, while the question of whether a particular norm or institution is a good means to such ends is mainly an empirical matter. In other words, political philosophy just tells that good politics provides a good answer to the question "what shall we do...?" and that the goodness of the answer is measured up against the ultimate ends of each political community. We are inclined to believe that such ultimate ends are universal as far as human beings are concerned (although this is a disputable empirical matter). However, their precise interpretation and ranking would vary historically and culturally.

Consequently, the political perspective, as understood by political minimalism, is not perniciously relativistic. Pernicious relativism in political philosophy would be the view that there are no objective criteria to distinguish good politics from bad politics. In turn, a valuable form of relativism is the view that what is a good political decision in one context may be a bad political decision in another context. According to the political minimalist, politics is an exercise of instrumental rationality aimed at realizing the ends of each political community. Although the ranking of ends may change from community to community and in different periods of each community's history, the constitutive aim of politics remains, and it is a matter of serving such ends. This constitutive aim provides a non-relativistic criterion for telling good politics from bad politics: this is a matter of whether or not, for each community at a certain occasion, its response to "what shall we do…?" is objectively good or not.

Secondly, in our view, the different realms of practical reason (i.e., prudential, moral and political) arise because we can pose different questions as regards the overall question what to do (i.e., "what shall I do as regards *my* ends and the means at my disposal?" vs. "what shall I do as regards *other people's* interests as ends of their own and not mere means to my interests?" vs. "what shall we do as regards *our* ends and the means at our disposal?"). This means that when we make a judgment on what to do, we do it as a response to a particular question. Yet, any decision or action can be assessed from either a prudential, moral, or political perspective. That is, we can assess whether such a decision or action is a good response to a prudential, moral, or political question.

Concerning the relationship between ethics and politics, political minimalism adopts a conception that is analogous to Quine's notion of reciprocal containment of epistemology and ontology (Quine 1969, p. 83), and claims that politics is prior to morality from the political point of view, while morality is prior to politics from the moral point of view (Author1 2017a, p. 730). So, contrary to what political moralists believe, in some contexts, political reasoning may vindicate its precedence over moral criteria. These are the contexts in which the most relevant question is not "what should I do?", but "what shall *we* do?"

As regards the topic of this paper, our thesis is not that we cannot address the question of how AVs should behave from a moral point of view, but only that, typically, such question arises from a political concern, and that, since there is a specific type of political normativity, it is possible to offer good political responses to it.

3. The limits of the analogy between AVs and moral agents

Contrary to what political moralists believe, we have argued that there are genuinely political questions whose answer ethics cannot provide. Moral questions, of course, continue to make sense in many contexts. Our aim in the rest of this article is to argue that the most relevant questions concerning the *regulation* of AVs are political instead of moral. As pointed out before, this does not mean that we cannot assess political responses from a moral perspective, and moral responses from a political perspective; only that neither the political nor the moral have always the precedence.

We started by describing an emergency in which a human driver had to choose whether to run over a group of pedestrians or to put at risk the travelers inside the vehicle that he or she was driving. Although it is not a purpose of this article to characterize morality and other normative practices outside politics, it is relatively uncontroversial, we think, to assert that the driver could, when making the decision, adopt one of these two perspectives: the perspective of what ought to be done, taking into account his or her interest (what is customarily called a "prudential" perspective) and the perspective of what ought to be done, taking into account others' interests or needs as ends in themselves and not only means to the agent's interests (the "moral" perspective). At any rate, it would not make sense to ask the driver to adopt on the spot a political perspective. Individual drivers who have to take such a kind of decision in a hurry do not face the characteristically political question "what shall we do...?" We could talk of a political deliberation if, for instance, a community of travelers was about to settle the policy of the group if such situations were to occur. Yet, even in such a peculiar setting, should the driver decide what to do in face of an emergency, this would still be either a moral or a prudential decision on whether or not to obey the instructions previously agreed by the group.

Importantly, we do not intend to suggest that political questions need to be made or answered by groups. We only mean that they are political questions inasmuch the corresponding responses can only be justified by adducing as reasons the ends of the group and the means at their disposal. For instance, questions concerning how the Law ought to regulate situations like the one described above, or how members of a community ought to be educated to react to them, are political questions, irrespective of how many people (perhaps only one solitary thinker) try to answer them in a given context.

A human driver who is compelled to react in an emergency faces either a prudential or moral concern. The literature on the ethics of AVs takes for granted the analogy between situations involving an individual human driver and other similar ones

involving AVs. Projects like *Moral Machine*, and most academic discussions on the topic, take the following analogy for granted: as far as decision-making is concerned, an AV is like a human driver, and it ought to do whatever a human driver ought to do in situations like the ones described above. As the *News Service of the University of Stanford* (2016) puts it, the aim is to "teach human ethics to autonomous cars". For this purpose, we first need to learn how human drivers would react (this is the information that MIT's *Moral Machine* and similar projects are supposed to provide) and then design the algorithms governing the behavior of cars accordingly. Then, software for AVs becomes another field in which building an ethics of algorithms is perceived as a pressing need (Ananny 2016; Bonnefon, Shariff, and Rahwan 2019).

But is the analogy adequate as a starting point for political decisions and legislation? One first consideration is that, although the attribution of rights and moral obligations to robots is a recurring theme in science fiction literature and the philosophy of robotics, AVs (at least, as regards the characteristics that they are expected to exhibit over the coming years) are far from being the kind of quasi-human entities that have stimulated literary imagination and philosophical speculation. In a letter addressed to the European Commission, the Signatories (2017) (a group of experts in robotics and artificial intelligence) complain that the current and foreseeable capacities of even the most advanced robots have been exaggerated as a result of a superficial understanding of the ability to self-learning of future robots. In these experts' opinion, this misunderstanding of the nature of robots would have led to the hasty introduction of notions such as the "legal status of the electronic person" by the European Parliament (2017). To date, robots are not moral agents, and it is doubtful that they will ever be. Consequently, AVs do not face moral dilemmas in the sense of wondering "What shall I do as regards...?", because, as the experts' letter says, they cannot really wonder anything as of yet, but only respond as trained.

Certainly, it is not *impossible* that robots could ever be moral agents. The concept of artificial moral agency appears more plausible if (i) we distinguish between moral agency and moral patienthood (it might be possible to recognize moral agency to robots before recognizing them rights, if ever), and (ii) we admit the existence of different levels of autonomy and, hence, of moral agency.[11] We can imagine, then, the possibility of robots that are similar enough to humans, as far as their level of autonomy is concerned, and that this is sufficient to grant them moral agency. Still, human societies would have to decide whether to allow or forbid the existence of such robots. For instance, van Wynsberghe and Robbins (2019) argue against the development of artificial moral agents in general. For the purposes of this article, we need to wonder only what sense would make to develop artificial moral *vehicles*. In the context of AVs, one should refrain from thinking of replicants (as those from the film *Blade Runner*), android hosts (as those from the series *Westworld*) or artificial doctors built with a human appearance to better fulfill their tasks. We ought to even stop thinking of an *individual* (a vehicle) taking decisions: each AV is rather an element in a transport system whose parts (AVs and others) are constantly

11 We are very thankful to one anonymous reviewer for valuable suggestions on this point.

interconnected and take decisions jointly (Roff 2018). For this reason, we prefer to read "AVs" as "automated vehicles" instead of "autonomous vehicles". On the other hand, these transport systems including AVs are designed, in contrast with human agents, to perform a single task: to carry goods and persons who, very importantly, are not their relatives, their friends, or themselves. It is hard to imagine what reasons a society might have to convert AVs into a sort of replicants with wheels and with the capacity for moral reasoning. Rather, societies will have good reasons to impose strict restrictions on the design of software for AVs. Lawmakers establishing such restrictions would be, consciously or not, answering the political question "what shall we do...?", concerning policies for the regulation of AVs and traffic.

Our insistence on legal restrictions does not entail, though, that lawmakers or software engineers can straightforwardly determine how AVs will respond in every situation they might encounter in real life. As Roff (2018) reminds us, transport systems to which AVs belong are learning systems informed by their experience and constrained by their architecture. These systems will have to make trade-offs between several values when producing an answer to novel traffic situations. But which values are those, together with other constrictions, will be given by political decisions translated into the laws that regulate software design and traffic norms.[12]

4. More problems with the use of ethics to regulate AVs behavior

Adopting ethics to establish the rules that AVs should obey both in emergency and other situations is a misled strategy for at least two more reasons. These further reasons are not specific for AVs since they equally hold as regards the legal regulation of the behavior of human drivers, but they show that many relevant questions commonly attributed to the ethics of AVs are political.

4.1. Societies, traffic, and the law

Political communities look at the behavior of drivers mainly because societies want to prevent casualties, injuries, and other damages and problems caused by traffic. The law forbids actions that ostensibly run counter to these aims, while other behaviors that are merely advisable are encouraged through traffic public policies. The considerations guiding laws and policies are, then, instrumental to achieve certain social ends. Consequently, moral considerations remain irrelevant when authorities judge the behavior of human drivers. Drivers are punished when their way of driving breaks the law, not when it is immoral (although in many cases we may agree that one particular conduct is both illegal and immoral). The same goes for carmakers or software programmers. According to political minimalism (and in this respect, we think political realists would agree), moral considerations are relevant for lawmakers, *as lawmakers*, only in the following way: the moral values that are present in a political community, and their relative weight, constitute a

12 As an anonymous reviewer has reminded us, legislation can be seen as a socially designed technology. This technology, in turn, partially determines the design of other technologies, like transport systems involving AVs.

relevant *fact* to consider when designing viable strategies to pursue the ends of the political community. In other words: from the political point of view, moral facts lose their straightforward normative force and become *social facts* with only indirect and partial normative weight.

The political perspective is thus at stake in the design of public policies and the enactment of laws, and this is the case when we discuss what AVs ought to do in emergencies. It is surprising, though, that the moral perspective has prevailed so far over the political in the literature concerning the behavior of AVs. Generally speaking, it is striking that the attention that is paid to the *ethics* of algorithms apparently supersedes the interest in the *politics* of algorithms (although in some cases the explanation may be that what is sometimes called "ethics" *is* in fact politics). We are claiming that a change in focus is needed, and also that important practical consequences follow from this change.

Against this view, Goodall (2014) argues that laws are not comprehensive or specific enough to cover every emergency that AVs can encounter, and he concludes that it is necessary to develop moral algorithms for AVs. Lin (2013) illustrates Goodall's point with this example: an AV governed by software that obeys the law would come to a full stop when meeting a small tree branch, instead of drifting into the opposite lane for a few seconds. This is so because the AV would dutifully observe traffic laws that prohibit crossing a double yellow line, while a human driver might make an exception if there is no oncoming traffic. The "ethical" human decision would be better than the "legal" machine choice because the latter might cause a crash with other vehicles.

Nevertheless, there is nothing in Goodall's argument or Lin's example that justifies the jump from the legal (and political) level to the moral one. The impossibility to foresee every possible situation does not cancel legal reasoning in favor of moral reasoning. Very often, in many different contexts, judges have to apply the law to novel contexts, but the argumentation they develop still responds to a legal, not moral, rationale. In many jurisdictions, a violation of the law —even a criminal law— is legally defensible under the traditional concept of necessity (*necessitas non habet legem*) that covers situations where an illegal action is the lesser of two evils, that is: "if the harm which will result from compliance with the law is greater than that which will result from violation of it" (LaFave and Scott 1986, §10.1). For instance, according to the Judicial Council of California (2017), it is justified to break the law to prevent significant bodily harm or evil when there was no adequate legal alternative, when the act did not create a greater danger than the one avoided, and when there was a reasonable belief that the act was necessary. In Lin's example, the decision to cross a double yellow line to prevent a crash with vehicles running behind, if there were no incoming traffic, would be justifiable on *legal* grounds.

Santoni de Sio (2017) has discussed the relevance of the legal doctrine of necessity and other legal principles for the programming of AVs to face emergencies. We think that this strategy is more promising than invoking ethics. One may long for a more straightforward link between the law and its particular applications, but the link

between particular cases and moral rules may be equally troublesome, and the situation is made even worse by the following reason: Appealing to ethical criteria makes us wonder what such criteria would be.

4.2. What moral criteria?

In contemporary societies, many moral codes and many rival theories in normative ethics coexist: which one should we choose as the theoretical framework in the case of the software for AVs? (Maxmen 2018). It is telling that some discussions on the ethics of algorithms, like many debates in other fields within applied ethics, begin with a description of the main normative ethical theories (Ananny 2016). Deontological approaches, utilitarianism, and virtue ethics are customarily mentioned, and then authors proceed either to take one of these frameworks for granted or to point out the consequences of choosing each of the listed theories for the discussed topic. Another possibility is appealing to intuitions. For instance, we could adopt the answers given by the majority to the *Moral Machine* tests worldwide. Or perhaps we should ponder the answers given by the majority in each sovereign state. Yet another possibility is resorting to professional codes of conduct that would reflect the consensus on the matter of a given profession, e.g., taxi drivers or engineers (Dennis, Fisher, Slavkovik, and Webster 2016). But appealing to intuitions or professional codes of ethics in the described ways would amount to committing the is/ought fallacy classically denounced by Hume (1739/2000): being the case that most people in a community, or most professionals, agree on what a good answer to a given moral question is does not entail that such answer is morally good indeed.[13] We could turn our eyes again to philosophers still debating the relative advantages and shortcomings of deontology, utilitarianism, and virtue ethics. Suppose we are not defenders of moral relativism (as a matter of fact, the authors of this article are not); suppose we think that there is a roughly correct moral theory and true sentences on topics such as how AVs ought to react to different traffic situations; suppose that we discover these truths, and suppose that we inform policymakers and legislators about the relevant true moral sentences on the behavior of AVs, backed by the correct moral theory. For instance, JafariNaimi (2018) points out the limitations of experimental ethics developed within a utilitarian framework and proposes to consider relevant contributions of *care ethics* and *situated knowledge*. Let us suppose, for the sake of argument, that this author is right and that we should build the ethics of algorithms on foundations that diverge from the most usual ones. Suppose we derive our conclusions on the ethics of AVs from this novel ethics of algorithms. Still, policymakers and legislators would

13 As Roff (2018) puts it in the context of AVs: "We can model moral dilemmas and ask people to partake in experiments, but that only tells us the empirical reality of what those people think. And that may be a significantly different answer than what morality dictates one ought do." Professional codes of conduct pose an additional problem: since they tend to reflect moral views that are not controversial, they do not provide substantial orientation in the face of dilemmas (Author1 2017b), so, by themselves, they would not be very useful for software programmers who are developing algorithms for AVs.

continue to face a problem: very likely, some "experts" in moral theory would provide advice on the matter that is incompatible with our expert advice; and handling this diversity of verdicts among experts (similar in degree, albeit not in sophistication, to the diversity of verdicts among laypeople) is a genuinely *political* problem, not a moral one:

> Both laypeople 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 policymakers any clear indication for addressing this issue (Santoni de Sio 2017).

Unfortunately, this author, although searching for the guidance of legal principles, continues to see "this issue" as a moral issue. We claim that it is more fruitful to address it as a political problem.

5. A political approach

Earlier, we introduced a conception of the relationship between ethics and politics that we call "political minimalism". According to this conception, politics is an attempt to answer the question "What shall *we* do as regards our ends and the means available to achieve them?" In other words, politics is an exercise of collective instrumental rationality aimed at accomplishing the ends of a given political community. We have claimed that, although we can judge political actions and institutions from a moral perspective, we can also look at moral criteria from a political perspective, and, when we do so, such criteria become social facts. We claim that as regards the regulation on AVs, the political perspective usually takes, and ought to take, precedence.

Let us go back to our initial example. Should a vehicle with a sudden brake failure run over a group of pedestrians, or should it swerve towards a wall? If the driver at stake is a human driver, the question might be answered from three different levels of practical reasoning. Firstly, the driver could adopt a *prudential* perspective, taking her or his interest and that of her or his family as the main concern. From this perspective, running over pedestrians is very likely to be a good decision. Secondly, the driver might take a *moral* stance. It is less clear what the driver ought to do, and what actions would at least be forgivable in this case, because the answer depends on what moral theory one accepts. Whether the driver decided on prudential or moral grounds, any of us feels legitimated to judge her or his decision from the moral point of view. Yet, society might want to impose a *political* response to this situation by developing legislation that establishes, for instance, that drivers killing pedestrians must be punished, but at the same time concedes that being at risk of death—be it oneself or one's family—would be a mitigating circumstance (together perhaps with other mitigating circumstances, like the uncertainty of the consequences of each course of action and the short time available to decide what to do).

While in the case of a human driver it is a disputed issue which perspective is more relevant, we claim that in the case of AVs only one is relevant to these settings: the political perspective. The reasons that motivate a human driver in such situations not to swerve to avoid harming her or his own family might prudentially and even morally justify her or his decision; yet, these reasons are absent in the case of an engineer developing software for AVs.

The lawmaker producing legislation on software for AVs is not considering a particular situation, let alone a dilemma involving her or his family. And precisely because of the generality of the problem, a political approach provides a better solution. Some precisions are needed, though, to explain what we understand by "generality" here.[14] The generality of the perspective does not have to do with the number of people affected by the decisions: there are moral decisions that affect many people (like the decision made by a wealthy person to donate to an NGO that fights famines), and political decisions that can affect only one person, like the decision by a government to free a prisoner. The type of generality relevant here is not a matter of the number of people who take the decision either: as we have already said, both reflection and political decision can be made by any number of people, including just one. The generality that is relevant here has to do with the adoption of the perspective of the community; it has to do with the attempt to answer the question that the community would hypothetically ask to itself about what it should do, regarding its policies on AVs, to achieve its ends. Then, lawmakers, who are expected to adopt this perspective, produce an answer that they intend to impose, through legislation, to software engineers, automakers, and users. If a lawmaker were to adopt a moral point of view, he or she would be asking him or herself what he or she ought to do concerning others (no matter these are few or many), not what the political community ought to do to promote its ends.

These ends that ultimately ought to guide political judgment are, in turn, the general ends of the community, not other, more contextual ends like the present priorities of the members of the community concerning AVs. In this respect, we diverge from Himmelreich (2018) and his understanding of the politics of AVs in terms of social choice. We shall return shortly on the difference between our approach and a social choice one.

Considering the general ends of the community, political judgment might lead to the conclusion that a reasonable goal at a lower level of generality is to save as many lives as possible through traffic policies, and that this goal is better achieved by protecting pedestrians in most circumstances because they are the weaker side. This is not a choice that favors one part of society to the detriment of another, for all of us can walk one day and travel by car the following. So, parliaments would be legitimated to enforce by law that AV software gives priority to the safety of pedestrians in all the cases similar to the one described above, and in most circumstances where the safety of pedestrians is at stake. A parliament approving

14 We are thankful again to one anonymous reviewer for urging us to be more explicit on this issue.

such a law would be doing it for political reasons and not for moral ones, according to our definition, even if moral considerations might be invoked—among other reasons, because preserving moral convictions as much and as far as possible is part of society's ends.

To say that we can dispense with the prudential and the moral perspectives when legislating on the behavior of AVs is not the same as saying that political answers are easy, nor that all societies are entitled to the same solutions to situations like the one that we are discussing: a given political community might decide, for instance, to rank individual freedom much higher than the security of its members. This is not a very peculiar choice—we know certain political communities that have adopted precisely this scale of values concerning guns.[15] Some may criticize this ranking of priorities as morally wrong, and others view it as morally irreproachable. But the legislators' job is not to undertake a moral debate, but to display the best means to accomplish the ends of their political community. Of course, there is an additional problem concerning how to discover what such ends are, how they are ranked in a particular community and how to better prioritize them to try to achieve as much of them as possible. We are not addressing this complex issue here, and it does not belong to the content of a philosophical proposal like political minimalism, because the mentioned problem is mainly an empirical matter. But political minimalism provides a criterion for judging different procedures for assessing this issue, like different kinds of democratic systems.

6. Practical consequences

A practical outcome of the previous considerations has been advanced already. When legislators in each country regulate the algorithms by which AVs will react to different traffic situations, they should not be searching for the morally correct answer, nor do they ought to try to respond to the moral intuitions of the majority or to the advice of philosophers, but instead, they should try to legislate to best serve the ends of the political community.

Himmelreich (2018, pp. 681–682) makes similar claims when he states that focus on mundane situations instead of trolley cases calls for "a social choice (a political solution)", and that "with autonomous vehicles, behavior in mundane situations becomes a matter of policy". But we would amend Himmelreich's claims in two respects. First, we are persuaded that not only ordinary mundane situations but also emergencies (more clearly once they are not treated as trolley cases) call for political solutions. One sound argument advanced by Himmelreich (2018, p. 678) to support his view that the driving behavior of AVs in mundane situations is a matter of general policy is the *difference of scale* with the case of a decision taken by a human driver: each individual does not feel the need to reflect in advance on his/her precise behavior in each mundane traffic situation because the way he or she drives does not make a significant difference overall. The difference is bigger in the case of

15 Some defenders of the freedom of citizens to own and use guns might say that their main concern is security. But since statistics show that weapons are a bad means to that end, it would make more sense for them to invoke freedom rather than security.

an algorithm intended to regulate the behavior of thousands of AVs.[16] But this is true both for ordinary situations and for emergencies, so the call for a political approach to AVs' algorithms is for the regulation of the whole behavior of AVs.

The second amendment concerns Himmelreich's mention of "social choice". As Keeling (2020, p. 304) rightly points out, both in a narrow and in a wide interpretation of the term, an appeal to social choice would entail that "the preferences, tastes or values of all the individuals in society uniquely determine the collective judgment". This conclusion poses the practical problem of deciding how to translate preferences, tastes, and values of all the individuals in a political community to every piece of legislation on algorithms for AVs. The political approach to AVs that we derive from political minimalism avoids this problem. According to our approach, preferences, tastes, and values of the members of the political community only have to be considered by legislators as relevant social facts, not as the criteria for the morally or politically correct solution to the question concerning the correct regulation of AVs' algorithms. The ultimate criteria would be provided not by individual preferences, tastes, and values, but by the general ends pursued by the political community as a whole. These are not a mere sum of individual preferences but an aggregate that involves a plan of prioritization to make them maximally compatible and reachable. Although we have acknowledged that finding out what those ends are and how they are ranked against each other in a given political community is not an easy task, it seems both more operational than asking the same question to every individual at every step of the political life, and more akin to what political practice is.

A second practical implication concerns the design of transportation systems. If we set aside the idea that an AV is a moral actor, it may be more difficult to assume that each model of AV (or each AV!) will use its own moral algorithm, a suggestion that is received with alarm by many (see Bonnefon et al. 2016). More interestingly, it might be easier to resist the idea that buyers could choose between different, more or less "altruistic" algorithms, perhaps by paying an extra when buying or renting an AV. It is true that not all moral stances would foster individualistic approaches to this debate and that not every political stance would be contrary to them. But if the debate is framed as a debate between moral theories or moral codes, as political moralists would want, a moral position that is more prone than others to fostering the *common* good (e.g., a utilitarian stance) would be just one of the many competing moralities available to the relevant political actors; and the very setting of a plurality of moralities plus the moralistic propensity of taking morality as the ultimate guidance would foster by itself the tendency to leave the selection of the moral algorithm in the hands of the buyer or user. The adoption of shared algorithms for

16 Note that we are contrasting the relatively little impact of a decision taken by an individual driver and the normally bigger impact of a decision taken by a legislative body, not two decisions taken by a legislative body, one concerning the regulation of AVs and the other concerning the regulation of the behavior of human drivers. If a human driving behavior is likely to have an important social impact, e.g., because it can be repeated by many drivers, legislators could feel the same need to approach that conduct politically and legislate accordingly, as they could have felt in the case of AVs.

the maximization of the goals of the community is more likely if society adopts a political perspective, although not all political stances will guarantee this maximization. A libertarian outlook would not guarantee it indeed, but most societies are not libertarian concerning traffic: laws in most countries impose speed limits, strict priority rules in crossroads, and the use of safety belts. Although political minimalism cannot discard by itself the possibility of a libertarian society that decides to rank individual freedom and self-ownership well above any other value, libertarianism regarding traffic laws would make sense more easily from an extremely moralistic, strong prioritization of individual moral rights, and would bring about very counter-intuitive consequences. So, again, the main problem might lie in political moralism and not just in the choosing of a particular moral code.

Transportation in the future might prioritize "automated highways" in which AVs are managed and controlled as part of a system, and hence favor the coordination of traffic according to general environmental and safety criteria, instead of the coexistence of individual vehicles weakly coordinated and responding to heterogeneous programming priorities (Royal Academy of Engineering 2009). As Stilgoe (2018) rightly points out, "autonomous vehicles" is an expression as deceiving as "self-driving cars", since such vehicles can function only as part of a fleet with which they share information constantly. Vehicles belonging to the same fleet also share software and are subject to the same norms. Now, car companies might want to keep their respective fleets independent from the rest and subject them to their own "code of ethics". We do believe that society should resist this trend and treat each AV as part of a single fleet, composed by all the AVs circulating within the borders of a given territory under the rules imposed by a given political community (e.g., a national state or the European Union). This conclusion would be consistent with the constructivist thesis that technologies are not just isolated artifacts, but complex networks relating human actors, non-human beings, and processes in a social and legal environment (Latour 2005). Besides, we are convinced that it is easier to arrive at the same conclusion from the framework provided by political minimalism (and, perhaps, political realism) than if political moralism is given for granted. We agree again with Stilgoe (2018) when he describes the emergence of AVs as "a process of social learning", and with JafariNaimi (2018), who points out that AVs offer an opportunity for reframing transportation governance through the socialization of machine learning.[17] As key elements in processes of social learning, AVs ought to be put under constant public scrutiny, and we believe that keeping in mind the difference between a political and a moral perspective would reinforce the rights of the political community to establish the necessary norms for the governance of AVs.

Finally, we have some comments concerning trolley cases and experimental ethics programs like *Moral Machine*. Are such strategies useful?

17 Winfield and Jirotka (2017) defend the inclusion of "ethical black boxes" in robotic systems to improve social learning. The idea fits our proposal well, although we would remove the word "ethical" to avoid confusion.

Although it is not our concern in this article, we might grant Keeling (2020) that the study of AVs is relevant for the *ethics* of AVs, and also that the ethics of AVs is a field both legitimate and useful. What we are claiming here is that some situations that are usually addressed through the ethics of AVs should be treated, being political problems, from a political perspective.

As far as surveys of moral judgment are concerned, we admit that their outcomes can be also interesting, provided that we do not ask them for what they cannot deliver. They are not going to solve any moral dilemma by providing the correct answer (remember the is/ought fallacy); nor will they provide direct guidance for congresspersons. But they can provide useful information about the moral values and opinions of the members of the community that, together with further information concerning other social facts, may be considered by lawmakers when regulating AVs. As we said earlier, good politics is oriented to realize the general ends of the political community, not the more restricted goals and values that citizens may uphold concerning smaller areas of debate, like traffic policies. The empirical knowledge that surveys can provide on these specific topics can be relevant not for the specification of the political ends but for designing the best means to achieve them.

At any rate, the empirical evidence that may be more useful in the design of public policies is not information about the judgments on crash scenarios that individuals utter based on their moral views. If what is at stake is the design of public policies and the development of laws in accordance with the values of the community, we expect that it would be more useful to ask citizens to adopt a political perspective and reflect not on their particular values but on the values of their communities and the public good. For this reason, it could be interesting to resort to procedures like consensus conferences, which also favor the communication between experts and citizens and the generation of informed consensus through deliberation (Nielsen, Hansen, Skorupinski, Ingensiep, Baranzke, Lassen, and Sandoe 2006).

The refinement of participatory and information-gathering tools is as necessary in this field as in any other context where public policies and social appropriation of technologies are at stake. There are many issues around AVs and, in general, around AI and the presence of algorithms in society on which the political communities will have to deliberate in the future. Some of these issues are: who asked for the AVs, and what is the problem for which they are the solution; who will pay the expensive infrastructure necessary for the circulation of AVs; who will set the prices for the use of those vehicles; how monopolies will be avoided in the management of transport systems; who will benefit from the huge amounts of data produced by the vehicles in circulation, and who will have access to them; how transportation networks should be designed to reduce their environmental impact (Floridi 2019) and, finally, yes, how AVs ought to behave when faced with emergencies.

It is useful to place questions concerning AVs in the broader framework of the social design of technologies.[18] The market is one of these technologies, and as such can be socially designed in several ways, although sometimes there is a tendency to identify the market with the version of it produced within capitalism, and to see the capitalistic market as a sort of natural and inescapable reality to which individuals have to adapt with the help of morality. But markets can be re-designed using another social technology, legislation, and the design of further technologies goes well beyond the scope of individual choice.

The issue of the response of AVs in emergencies is, then, just an example of the political dimension of all technologies, and also of the dangerous tendency to believe that everything a society has to say about a specific technology can be reduced to individual moral judgments regarding the *use* of that technology, once it has been developed and is available in the market. Such belief is wrong for any technology, and it is even more dangerous if we intend that AVs can produce those moral judgments. We hope that this article provides one more argument to justify the political oversight that societies have the right to exercise over any technology from its outset. Technologies like artificial intelligence, the law, and the market are forces far too powerful to leave ethics as our only tool for their social design.

18 We thank one anonymous reviewer for the suggestion that it might be common to technologies that they are all designed. Additionally, we characterize technologies as the result of a social "shaping", echoing the usual idea of the social construction of technologies, although not committing ourselves to a straightforward social determinism.

REFERENCES

Ananny, M. (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology & Human Values*, 41(1), 93–117.

Author1 (2017a).

Author1 (2017b).

Author2 (2011).

Author2 & Author1 (forthcoming).

Author2 & N. Villanueva (forthcoming).

Awad, E., Dsouza. S., Kim, R., Schulz, J., Henrich J., Shariff A., Bonnefon J.F., & Rahwan I. (2018). The Moral Machine experiment. *Nature*, 563, 59–64.

Bonnefon, J.F., Shariff, A., & Rahwan, I. (2016). The Social Dilemma of Autonomous Vehicles. *Science*, 352, 1573–76.

Bonnefon, J., Shariff, A., & Rahwan, I. (2019). The Trolley, The Bull Bar, and Why Engineers Should Care About The Ethics of Autonomous Cars. *Proceedings of the IEEE*, 107, 502–504.

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 99, 906–944.

Cohen, G. A. (2008). *Rescuing Justice and Equality*. Cambridge, Mass.: Harvard U.P.

Deng, B. (2015). Machine ethics: The robot's dilemma. *Nature*, 523, 24–26.

Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal Verification of Ethical Choices in Autonomous Systems. *Robotics and Autonomous Systems*, 77, 1–14.

Erman, E., & Möller, N. (2013). Political Legitimacy in the Real Normative World: The Priority of Morality and the Autonomy of the Political. *British Journal of Political Science*, 45, 215–233.

Erman, E., & Möller, N. (2015). Why Political Realists Should Not Be Afraid of Moral Values. *Journal of Philosophical Research*, 40, 459–464.

Etzioni, A., & Etzioni, O. (2016). AI Assisted Ethics. *Ethics and Information Technology*, 18(2), 149–56.

European Parliament (2017). *Civil Law Rules of Robotics Resolution*. http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html?redirect. Accessed 25 April 2020.

Floridi, L. (2019). Autonomous Vehicles: from Whether and When to Where and How. *Philosophy & Technology*, 32, 569–573.

Foot, P. (2002). *Virtues and Vices*. Oxford: Oxford University Press.

Galston, W. (2010). Realism in Political Theory. *European Journal of Political Theory*, 9, 385–411.

Gerdes, J.C., & Thornton, S.M. (2015). Implementable Ethics for Autonomous Vehicles. In M. Maurer, J.C. Gerdes, B. Lenz & H. Winner (eds), *Autonomes Fahren* (pp. 87–102). Berlin: Springer.

Geuss, R. (2008). *Philosophy and Real Politics*. Princeton, NJ: Princeton University Press.

Gogoll, J., & Müller. J.F. (2017). Autonomous Cars: In Favor of a Mandatory Ethics Setting. *Science and Engineering Ethics*, 23, 681–700.

Goodall, N.J. (2014). Machine Ethics and Automated Vehicles. In G. Meyer & S. Beiker (eds), *Road Vehicle Automation* (pp. 93–102). Cham: Springer.

Goodall, N.J. (2016). Can you program ethics into a self-driving car? *IEEE Spectrum*, 53(6), 28–58.

Greene, J.D. (2016). Our Driverless Dilemma. *Science*, 352, 1514–15.

Greenemeier, L. (2016). Driverless Cars Will Face Moral Dilemmas. *Scientific American*, June 23, 2016. https://www.scientificamerican.com/article/driverless-cars-will-face-moral-dilemmas/ Accessed 19 July 2019.

Himmelreich, J. (2018). Never mind the trolley: The ethics of autonomous vehicles in mundane situations. *Ethical Theory and Moral Practice*, 21, 669–684.

Hume, D. (1739/2000). *A Treatise of Human Nature*. Edited by Norton DF, Norton MJ. Oxford Univ. Press, Oxford, III, I, I.

JafariNaimi, N. (2018). Our Bodies in the Trolley's Path, or Why Self-driving Cars Must *Not* Be Programmed to Kill. *Science, Technology, & Human Values*, 43(2), 302–323.

Jubb, R. (2019). On What a Distinctively Political Normativity Is. *Political Studies Review*, 17, 360–369.

Jubb, R., & Rossi, E. (2015). Political Norms and Moral Values. *Journal of Philosophical Research*, 40, 455–458.

Judicial Council of California (2017). *California Criminal Jury Instructions* §3403. https://www.justia.com/criminal/docs/calcrim/3400/3403/. Accessed 20 July 2019.

Kant, I. (1781/1998). *Critique of Pure Reason*. Translated and edited by P. Guyer and A.W. Wood. Cambridge: Cambridge University Press.

Kant, I. (1795/1970). *Perpetual Peace: A Philosophical Sketch*. Translated by H. B. Nisbet. In H.S. Reiss (ed), *Kant: Political Writings*. Cambridge: Cambridge University Press.

Keeling, G. (2020). Why Trolley Problems Matter for the Ethics of Automated Vehicles. *Science and Engineering Ethics*, 26, 293–307.

Kirkpatrick, K. (2015). The Moral Challenges of Driverless Cars. *Communications of the ACM*, 58(8), 19-20.

LaFave, W.R., & Scott, A.W. (1986). *Substantive criminal law*. St. Paul, Minn: West Pub. Co.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford UP.

Larmore, C. (2013). What Is Political Philosophy? *Journal of Moral Philosophy*, 10 (3), 276-306.

Leader Maynard J., & Worsnip, A. (2018). Is There a Distinctively Political Normativity? *Ethics*, 128, 756–787.

Lin, P. (2013). The Ethics of Autonomous Cars. *The Atlantic*, 8 October 2013. https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/ Accessed 19 July 2019.

Lin, P. (2016). Why Ethics Matters for Autonomous Cars. In M. Maurer, J.C. Gerdes, B. Lenz & H. Winner (eds), *Autonomes Fahren* (pp. 69–85). Berlin: Springer.

Maxmen, A. (2018). Self-driving car dilemmas reveal that moral choices are not universal. *Nature*, 562, 469-470.

McBride, N. (2016). The Ethics of Driverless Cars. *ACM SIGCAS Computers and Society*, 45(3), 179–84.

Moor, J.H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), 18–21.

Newey, G. (2010). Two dogmas of liberalism. *European Journal of Political Theory*, 9 (4), 449–465.

News Service of the University of Stanford (2016). Stanford researchers teach human ethics to autonomous cars https://news.stanford.edu/2016/08/01/teach-autonomous-cars-drive-like-humans/. Accessed 19 July 2019.

Nielsen, A.P., Hansen, J., Skorupinski, B., Ingensiep, H.W., Baranzke, H., Lassen, J., & Sandoe, P. (2006). *Consensus Conference Manual*. The Hague: LEI.

Nyholm, S., & Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? *Ethical Theory and Moral Practice*, 19, 1275–1289.

Purves, D., Jenkins, R., & Strawser, B.J. (2015). Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice*, 18 (4), 851–72.

Quine, W. V. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Roff, H. (2018). The folly of trolleys: Ethical challenges and autonomous vehicles. *Brookings*. https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/ Accessed 25th April 2020.

Rossi, E. (2012). Justice, legitimacy, and (normative) authority for political realists. *Critical Review of International Social and Political Philosophy*, 15(2), 149–164.

Rossi, E. (2019). Being realistic and demanding the impossible. *Constellations*, 26, 638–652.

Rossi, E., & Sleat, M. (2014). Realism in Normative Political Theory. *Philosophy Compass* 9(10), 689–701.

Royal Academy of Engineering (2009). *Autonomous Systems: Social, Legal, and Ethical Issues.* London: Royal Academy of Engineering.

SAE (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_201806/ Accessed 25th April 2020.

Sangiovanni, A. (2008). Justice and the Priority of Politics to Morality. *The Journal of Political Philosophy*, 16, 137–164.

Santoni de Sio, F. (2017). Killing by Autonomous Vehicles and the Legal Doctrine of Necessity. *Ethical Theory and Moral Practice*, 20, 411–429.

Sleat, M. (2016). Realism, Liberalism, and Non-ideal Theory. Or, Are there Two Ways to do Realistic Political Theory? *Political Studies*, 64, 27–41.

Stilgoe, J. (2018). Machine learning, social learning, and the governance of self-driving cars. *Social Studies of Science*, 48(1), 25–56.

The signatories (2017). Robotics Open Letter. http://www.robotics-openletter.eu. Accessed 20 July 2019.

Walzer, M. (1983). *Spheres of Justice*. New York: Basic Books.

van Wynsberghe, A., & Robbins, S. (2019). Critiquing the Reasons for Making Artificial Moral Agents. *Science and Engineering Ethics*, 25, 719–735.

Weinberg, J. (2017). Philosophers Awarded Over $500,000 To Study Autonomous Vehicles. *Daily Nous, News for and about the Philosophy Profession*. http://dailynous.com/2017/10/06/philosophers-awarded-500000-study-autonomous-vehicles/. Accessed 19 July 2019.

Williams, B. (2005). *In the beginning was the deed: realism and moralism in political argument*. Princeton, N.J.: Princeton University Press.

Winfield, A.F.T., & Jirotka, M. (2017). The case for an ethical black box. In Y. Gao, S. Fallah, Y. Jin & C. Lekakou (eds), *18th Conference Towards Autonomous Robotic Systems* (pp. 262–273). London: Springer.