ACCIDENTS INVOLVING AUTONOMOUS VEHICLES: LEGAL ISSUES AND ETHICAL DILEMMAS

Giuseppe Contissa / Francesca Lagioia / Giovanni Sartor

Adjunct Professor & Research Assistant, CIRSFID – University of Bologna
Via Galliera 3, 40121, Bologna, IT
giuseppe.contissa@unibo.it; https://www.unibo.it/sitoweb/giuseppe.contissa

Research Assistant, Law Department, EUI – European University Institute
Via Bolognese 156, 50139 Florence, IT (Villa Salviati, Castello – Office SACA 104)
francesca.lagioia@eui.eu; http://www.eui.eu/Directory/Detail.aspx?userid=10273

Full Professor, CIRSFID – University of Bologna
Via Galliera 3, 40121, Bologna, IT
giovanni.sartor@unibo.it; https://www.unibo.it/sitoweb/giovanni.sartor

Abstract: *Accidents involving autonomous vehicles (AVs) raise difficult ethical dilemmas and legal issues. In this paper, we argue that self-driving cars should not be programmed to kill, that is, they should not be equipped with pre-programmed approaches to ethical dilemmas. On the contrary, we believe that AV systems should be designed in such a way that only user/passenger has the task (and burden) of deciding what ethical approach should be taken in unavoidable accident scenarios. We thus propose that AVs be equipped with what we call an «Ethical Knob», a device enabling the passenger to choose between different settings corresponding to different moral approaches or principles. An AV would accordingly be entrusted only with implementing the user's ethical choices, while the manufacturer/programmer would be tasked with enabling the user's choice and ensuring its implementation by the AV.*

1. Introduction

Some recent works have focused on the ethical dilemmas emerging from hypothetical accident scenarios where Autonomous Vehicles are entrusted with making decisions involving the lives of passengers and of third persons [Bonnefon et al. 2016, 2015; Nyholm/Smids 2016; Lin 2016]. In particular, the decisions an autonomous vehicle could make in the moments leading up to an impending collision have been framed by reasoning from the «trolley problem», a classic ethical thought experiment discussed by Foot [1967] and Thomson [1976].

In this contribution, we propose an approach to that restores to the passenger the task of choosing the ethical approach through which the AVs will address unavoidable accident scenarios. This solution we will call the «Ethical Knob».

In a recent paper, Bonnefon et al. [2016] state that accidents involving AVs create the need for new kinds of regulation, especially in cases where harm cannot be entirely avoided. In fact, while it is expected that AVs will generally reduce the number of traffic accidents, some accidents will still be unavoidable. In some occurrences of unavoidable accident, AVs must choose between running over pedestrians or sacrificing themselves and their passenger.

To illustrate the ethical and legal dilemmas raised by the use of AVs under such circumstances, Bonnefon et al. [2016] consider three scenarios involving imminent unavoidable harm:

– The AV can either stay on course and kill several pedestrians or swerve and kill one passerby.
– The AV can either stay on course and kill one pedestrian or swerve and kill its own passenger.

    – The AV can either stay on course and kill several pedestrians or swerve and kill its own passenger.

The common factor in all these scenarios is that the harm to persons is unavoidable, and so that a decision needs to be made as to which person will be harmed: the passenger, the pedestrian, or the passerby.

Such critical decisions, according to Bonnefon et al. [2016], ought to be delegated to the AV. This would require the manufacturer to develop a moral algorithm and to program the AV so that it can make the best decision. Bonnefon et al. [2016] also submit that this decision should be made on a utilitarian or consequentialist approach. Utilitarian AVs should be programmed to minimize the overall death toll, even if under some circumstances this may require sacrificing the life of the passenger(s). Therefore, in Bonnefon et al. [2016]'s approach, the life and death decision is made by the AV on the basis of the instructions provided by the manufacturer/programmer.

In this paper, we will argue for a different approach: AV systems should be designed in such a way that only the passenger has the task (and burden) of deciding what ethical approach should be taken in unavoidable accident scenarios. The machine should only be tasked to implement the user's ethical approach, based on its risk assessment. To make this point, we will first illustrate how life and death dilemmas are dealt with by the law, when a human is driving the car. We will then consider how they will be dealt when their solution is preprogrammed by the vehicle designer/manufacturer. And, finally, we will suggest how the ethical decision can be given back to the human driver, enabling him or her to choose the ethical approach that the AV will implement. We will also argue that this approach may avoid some of the perplexing issues involved in the ethical preprogramming of the AVs.

2. Liability analysis

In this section, we will analyse how the allocation of legal liabilities varies when the accident involves different kinds of vehicle: a man-driven car, an ethically preprogrammed AV, an AV whose ethical approach is selected by the passenger.

2.1. Man-driven car

Let us first assume that consider the car is driven by a human who did not contribute to creating the danger.

It seems to us that under this assumption, all three scenarios described in section 1, the choice to stay on course, which leads to the death of pedestrians, can be legally justified, so that the driver might avoid punishment.

In scenario (a), the choice to stay on course and let several pedestrians be killed, rather than to swerve and kill one passerby, can be justified on the moral-legal stance condemning the wilful causation of death (as distinguished by letting death result from one's omission).

In scenario (b), the choice to stay on course can be justified by invoking the state of necessity, since this choice is necessary to save the life of the driver.

The same justification applies to scenario (c), even though in this case the driver's choice to save his or her own life leads to the death of several other persons.

2.2. Preprogrammed AV

Let us now assume that car's the behaviour of the car has been preprogrammed.

We just saw that in scenario (a) the driver may be legally justified when choosing to stay on course and let several pedestrians be killed, rather than to swerve and kill one passerby.

In scenario (a) it is doubtful whether the programmer would be justified when choosing to program an AV so that it stays on course and kills several pedestrians rather than swerving and killing just one passerby. In fact, the distinction between omitting to intervene (letting the car follow its path) and acting in a determined way (choosing to swerve) – a distinction that in the case of a manned car may justify the human choice of allowing

the car to keep going straight, as we saw in Section 2.1 – does not seem to apply to the programmer, since the latter would deliberately choose to sacrifice a higher number of lives.

In scenario (b), is very doubtful whether preprogramming the car either to go straight (killing a pedestrian) a to swerve (killing the passenger) would be legally acceptable: in both case the programmer would arbitrarily choose between two lives.

In scenario (c), it seems that preprogramming the car to continue on its trajectory, causing the death of a higher number of people could not be morally or legally justified: it would amount to an arbitrary choice to kill many rather than one.

Our analysis of the three scenarios shows that some preprogrammed choices would be morally and legally unacceptable, even when the corresponding choices by the driver would be legally acceptable or at least excusable.

By further examining the legal and practical implication of pre-programming the car behaviour, some further inconvenient implications emerge. Let us assume that pre-programmed AV cars are introduced in a competitive market, without legal constraints on the choices just outlined. It seems to us that market pressures would encourage the introduction of AV cars programmed in such a way as to have a preference for the passenger's safety (considering that it is the passenger who would choose what car to buy or rent). This would put the lives of pedestrians at risk. The risk to pedestrians would substantially increase if cars were programmed in such a way as to always minimise the risk to the passenger, whatever risks this choice entails for third parties.

2.3. User-selectable ethical standard: the ethical knob

Let us now imagine that the AV is fitted with an additional control, the «Ethical Knob» (see Figure 1).
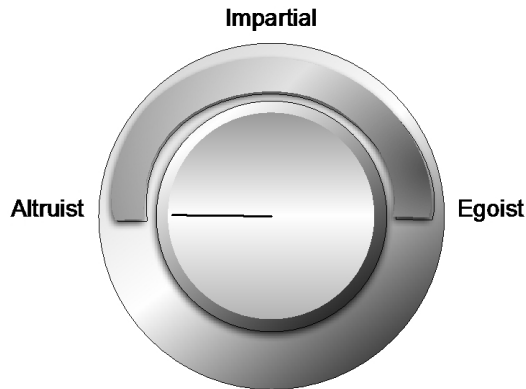

Figure 1: The Ethical Knob

The knob gives the passenger the option to select one three settings (see Figure 1):

1. **Altruistic Mode**: preference for third parties;
2. **Impartial Mode**: equal importance given to passenger(s) and third parties;
3. **Egoistic Mode**: preference for passenger(s).

In the first mode (altruistic), other people's lives outweigh the life of the AV passenger. Therefore, the AV should always sacrifice its own passenger(s) in order to save other persons (pedestrians or passersby).

In the second mode (impartial), the lives of AV passenger(s) stand on the same footing as the lives of other people. Therefore, the decision as to who is to be saved and who is to be sacrificed may be taken on utilitarian

grounds, e.g., choosing the option that minimises the number of deaths. In cases of perfect equilibrium (where the number of passengers is the same as that of third parties), there might be a presumption in favour of passengers, for the third parties, or even a random choice between the two.

In the third mode (egoistic), the passenger's life outweighs the lives of other people. Therefore, the AV car should act so as to sacrifice pedestrians or passersby rather than its own passenger.

The functioning of the knob, at least in principle, can be extended so as to include kin altruism, so that, in the Egoist mode, the AV will always act to save not only the passenger, but also his or her family or significant others.

Let us now assume that an AVs is endowed with the Ethical Knob.

The allocation of liability would be in principle be the same as for manned cars. However, since the car's behaviour has to be chosen beforehand, there should be no difference between omissive behaviour (letting the car proceed in its course) and active behaviour (swerving to avoid pedestrians on the street).

In scenario (a) the passenger's life is not at stake; therefore the setting of the knob does not matter. Consequently, the AV's behaviour should be based on utilitarian grounds: it should follow the trajectory that minimises the number of deaths. In fact, since the knob's setting is decided in advance relatively to the accident, a choice to keep going and kill several pedestrians rather than a single passerby cannot be justified according to a moral stance that condemns the active causation of death more than the omissive failure to prevent it.

In scenario (b) and (c), by contrast, the passenger's life *is* at stake; therefore the car's behaviour would depend on the setting of the knob. Moreover, since the passenger's life is directly at stake – and the passenger is aware of this possibility when setting the knob – the general state-of-necessity defence will apply, excusing the driver's choice to prioritise his or her life.

More specifically, in scenario (b) we could have the following behaviour, depending on the knob setting.

(1) If the knob is set to egoistic mode, the AV car will always act to sacrifice pedestrians or passersby in order to save its own passenger. (2) If the knob is set to impartial mode, the AV will take a utilitarian approach, thus minimising the number of deaths (and deciding according to a predefine default or randomly, when the number is the same for both choices). (3) If, finally, the knob is set to altruistic mode, the AV will sacrifice its own passenger in order to save pedestrians or passerby.

In scenario (c) the AV's behaviour will be the following. (1) If the AV is set to egoistic mode, it will always save its own passenger. (2, 3) In impartial mode, as well as in altruistic mode setting, the AV will sacrifice its own passenger in order to save several pedestrians.

In scenarios (b) and (c), the applicability of the state-of-necessity defence will exclude criminal liability, but the passenger could still be civilly liable for damages and be required to pay compensation. In this regard, the different knob settings presented above may affect third-party insurance. Presumably, the insurance premium will be higher if the passenger chooses to sacrifice other people's lives in order to save him/herself.

We have so far assumed that the knob has just three settings: egoism (preference for the passenger), impartiality, and altruism (preference for the third parties). These preferences are sufficient to determine a choice, assuming a deterministic context, i.e., that in every possible situation at hand it is certain what lives will be lost, whether the car keeps a straight course or swerves.

In real-life examples the situation may be much fuzzier: each choice (holding a straight course or swerving) may determine ex ante only a certain probability of harm (for the passenger or for a third party).

To address these situations we need a knob that allows for continuously changing settings, each one determining the weight of the life of the passenger(s) relative to that third parties.

Besides, in our scenario we have assumed that choices are between the life of a single passenger and that of a single third party. The model should be extended to cover cases where more than two lives are at stake.

3. Conclusions

The moral dilemmas where AVs must choose the lesser of two evils raise many ethical and legal issues. The Ethical Knob addresses these issues by giving back to the passenger the moral decisions and the judgement as to which outcome is more acceptable.

The Ethical Knob allows the passenger to choose between different settings corresponding to different moral approaches, i.e., general principles of conduct; the AV is only entrusted with implementing the user's choices, while the role of the manufacturer/programmer is to enable the different settings, and ensure their implementation into the AV, according to the user's choice. Therefore, with the Ethical Knob, the AV's decisions in the face of moral-legal dilemmas depends not on the designer but on the user.

Thus, the AV's moral dilemmas do not fundamentally differ from decision-making problems faced by human drivers on manned vehicles.

From a legal perspective, there could be different non-punishable choices when someone is faced with ethical dilemmas involving unavoidable deaths (or risks of death): the legal permissibility (or at least the non-punishability) of given choices will depend on the scope allowed for the state-of-necessity defence.

With the Ethical Knob, responsibility for ethical decisions would shift back to the users, and the state-of-necessity defence would work in some cases as it would for drivers in traditional cars. However, the fact that the setting on the knob is selected in advance can affect some contexts, particularly where the action/omission distinction could be applied to justify a non-consequentialist approach for a human driver.

Regarding AVs equipped with the Ethical Knob, in principle no obligations or liabilities other than those provided for manned vehicles would fall to the producer/programmer. This could facilitate the placement of AVs on the market.

Furthermore, the Ethical Knob may improve users' acceptance of AVs, giving users the ability to choose a moral algorithm that reflects their moral attitudes and convictions.

4. References

BONNEFON, J.-F./ SHARIFF, A./RAHWAN, I., Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars?, 2015. arXiv preprint arXiv:1510.03346.

BONNEFON, J.-F./ SHARIFF, A./RAHWAN, I., The social dilemma of autonomous vehicles, Science 2016, 352(6293), pp. 1573–1576.

FOOT, P., The problem of abortion and the doctrine of double effect, Oxford Review 1967, no. 5.

LIN, P., Why ethics matters for autonomous cars, Autonomous Driving 2016, pp. 69–85.

NYHOLM, S./SMIDS, J., The ethics of accident-algorithms for self-driving cars: an applied trolley problem?, Ethical Theory and Moral Practice 2016, pp. 1–15.

THOMSON, J. J., Killing, letting die, and the trolley problem, The Monist 1976, 59(2), pp. 204–217.