

*Extended Abstract*

# The Misconception of Ethical Dilemmas in Self-Driving Cars <sup>†</sup>

Tobias Holstein <sup>1,2,\*</sup>

<sup>1</sup> School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden

<sup>2</sup> Department of Computer Science, University of Applied Sciences, Darmstadt, Germany

\* Correspondence: tobias.holstein@mdh.se

† Presented at the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12-16 June 2017.

Published: date: 9 June 2017

**Abstract:** Self-driving cars are a transdisciplinary topic and well discussed in public and science. However, ethical dilemmas, such as the trolley problem, seem to dominate those discussions and consequently obfuscate much bigger ethical challenges in the development and operation of self-driving cars. We propose a systematic approach by creating a conceptual ethical model that connects components, systems and stakeholders to pinpoint ethical challenges for self-driving cars. This will help to move away from stagnating discussions over abstract thought experiments and to move forward to address and solve actual ethical challenges.

**Keywords:** Ethics; Self-Driving Car; Autonomous Car; Decision Making; Trolley Problem

---

## 1. Introduction

Self-driving, also called fully autonomous or driverless cars are in focus in many domains, such as engineering, computer science, human-computer interaction and ethics. From an engineering and scientific perspective, technical problems are challenging, but are solved one step at a time. When it comes to ethics, it seems that many discussions run into a dead end. In a constructed ethical dilemma there is per definition no solution: whatever you do, the result will be bad.

The trolley problem, which is an ethical thought experiment [1], is a commonly used example of an unsolvable ethical dilemma: The self-driving car drives on a street with high speed. A group of people suddenly appears in front of the car. The car is too fast to stop before it reaches the group. If the car does not react, the whole group will be killed. The self-driving car could however evade the group by entering the pedestrian way and consequently killing a previously uninvolved pedestrian (Option A). Replacing the pedestrian with a concrete wall, which in consequence will kill the passenger of the self-driving car, is another option (Option B). Varying the personas of people in the group, the single pedestrian or the passenger can be used to alter the experiment. The use of personas allows including an emotional perspective [2], such as, e.g., stating that the single pedestrian is a child, a relative, very old, very sick or a brutal dictator, who killed thousands of people.

Even though the scenarios are similar, the responses of humans, when asked how they would decide, differ [3]. The problem is that the question asked has a limited number of possible answers, which are all ethically questionable and perceived as bad or wrong. Therefore, a typical approach to this problem is to analyze the scenarios by following ethical theories, such as e.g., utilitarianism, other forms of consequentialism or deontological ethics. For example, utilitarianism would aim to minimize casualties, even if it means to kill the passenger, by following the principle: the moral action is the one that maximizes utility (or in this case minimize the damage). Depending on the doctrine, different arguments can be used to prove or disprove the decision.

Applying ethical doctrines to analyze a given dilemma and possible answers can only be done by humans. How would self-driving cars solve such dilemmas? There are a number of publications that suggest to implement moral principles into algorithms of self-driving cars [3]–[6]. We find that this does not solve the problem, but it reassures that the solution is calculated based on a given set of rules or other mechanisms, moving the problem to engineering, where it is implemented.

## 2. Ethical vs. Engineering Problems and Decision Making

It is worth to notice that the engineering problem is substantially different from the hypothetical ethical dilemma. While an ethical dilemma is an idealized constructed state that has no good solution, an engineering problem is always by construction such that it can differentiate between better and worse solutions. A decision making process that has to be implemented in a self-driving car can be summarized as follows. It starts with an awareness of the environment: Detecting obstacles, such as a group of humans, animals or buildings, and also the current context/situation of the car using external systems (GPS, maps, street signs, etc.) or locally available information (speed, direction, etc.). Various sensors have to be used to collect all required information. Gaining detailed information about obstacles would be a necessary step before a decision can be made that maximizes utility/minimizes damage. A computer program calculates solutions and chooses the solution with the optimal outcome. The self-driving car executes the calculated action and the process repeats itself.

## 3. Identifying Ethical Challenges

The process itself can be used to identify concrete ethical challenges within the decision making by considering the current state of the art of technology and its development. In a concrete car both the parts of this complex system and the way in which it is created have a critical impact on the decision-making. This includes for instance the quality of sensors, code and testing. We also see ethical challenges in design decisions, such as whether a certain technology is used because of its lower price, even though the quality of information for the decision making would be substantially increased if more expensive technology (such as sensors) would be used.

Since building and engineering of self-driving vehicle involves various stakeholders, such as software/hardware engineers, sales people, management, etc., we can also pose the following questions: does the actual self-driving car have a moral on its own or is it the moral of its creators? And who is to blame for the decision making of a self-driving car?

## 4. Approach

Prototypes of self-driving cars are already participating in public traffic among human-driven cars [7]. Therefore it is important to investigate how self-driving cars are actually built, how ethical challenges are addressed in their design, production and use, and how certain decisions are justified. Discussing this before self-driving cars are officially introduced into the market, allows taking part in the setting and definition of ethical ground rules. McBride states that “Issues concerning safety, ethical decision making and the setting of boundaries cannot be addressed without transparency” [8]. We think that transparency is only one factor, as it is necessary to start further investigations and discussions.

In order to give a more detailed perspective on the complex decision making process, we propose to create a conceptual ethical model that connects the different components, systems and stakeholders. It will show interdependencies and allow pinpointing ethical challenges.

Focusing on important ethical challenges that should currently be addressed and solved is an important step before ethical aspects of self-driving cars can actually be meaningfully discussed from the point of view of societal and individual stakeholders as well as designers and producers. It is important to focus not on abstract thought experiments but on concrete conditions that influence the behavior of self-driving cars and their safety as well as our expectations from them.

## References

- [1] P. Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Rev.*, vol. 5, 1967.
- [2] A. Bleske-Rechek, L. A. Nelson, J. P. Baker, M. W. Remiker, and S. J. Brandt, "Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a romantic partner," *J. Soc. Evol. Cult. Psychol.*, vol. 4, no. 3, pp. 115–127, 2010.
- [3] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science (80-. )*, vol. 352, no. 6293, pp. 1573–1576, 2016.
- [4] N. J. Goodall, "Can you program ethics into a self-driving car?," *IEEE Spectr.*, vol. 53, no. 6, 2016.
- [5] L. Dennis, M. Fisher, M. Slavkovik, and M. Webster, "Formal verification of ethical choices in autonomous systems," *Rob. Auton. Syst.*, vol. 77, pp. 1–14, 2016.
- [6] L. Dennis, M. Fisher, M. Slavkovik, and M. Webster, "Ethical choice in unforeseen circumstances," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8069 LNAI, pp. 433–445.
- [7] M. Persson and S. Elfström, "Volvo Car Group's first self-driving Autopilot cars test on public roads around Gothenburg," *Volvo Car Group Press Release*, 2014. [Online]. Available: <https://www.media.volvocars.com/global/en-gb/media/pressreleases/145619/volvo-car-groups-first-self-driving-autopilot-cars-test-on-public-roads-around-gothenburg>.
- [8] N. McBride, "The Ethics of Driverless Cars," *SIGCAS Comput. Soc.*, vol. 45, no. 3, pp. 179–184, Jan. 2016.



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)