



HAL
open science

Autonomous vehicles: Why customers should not buy a car programmed to kill?

Olivier Koumba

► **To cite this version:**

Olivier Koumba. Autonomous vehicles: Why customers should not buy a car programmed to kill?. 2025. hal-04890105

HAL Id: hal-04890105

<https://univ-angers.hal.science/hal-04890105v1>

Preprint submitted on 16 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autonomous vehicles: Why customers should not buy a car programmed to kill?

Olivier Koumba¹

Univ Angers, GRANEM, SFR CONFLUENCES, F-49000 Angers, France

ABSTRACT

The emergence of artificial intelligence raises questions about the moral dilemmas that the development of autonomous vehicles confronts us with. To answer these questions, many papers have proceeded to a quantitative description of moral facts and assume that autonomous vehicle is a moral agent in the strict meaning, i.e., autonomous vehicles will sometimes have to choose between two evils: running over pedestrians or sacrificing themselves and their passenger to save the pedestrians. Participants in a study conducted by Moral Machine (2018) approve of autonomous vehicles that sacrifice their passengers for the greater good and would like others to buy them, but they themselves would prefer to travel in autonomous vehicles that protect their passengers at all costs. Why? This research is the first that speaks to the business impact of the moral dilemmas of autonomous vehicle in an accident situation by focusing on the purpose of any vehicle: to serve all market stakeholders.

Keywords: Consumer behavior; Purchasing behavior; Moral dilemmas; Autonomous vehicle; Human behaviour; Moral judgments.

1 Introduction

Elaine Herzberg was struck and killed in 2018 in the first road accident involving an autonomous vehicle (McCausland, 2019). The vehicle was part of a test by the company Uber, which afterwards suspended its program. To offer an analogy, while a kitchen knife may be

¹ I would like to thank Renaud de Sainte-Marie; Victoria Teresa K., Victoria Voute K. who were always available to provide feedback and insightful comments at various stages of this document. I would also like to thank Guillaume de Tanoüarn for seeing some merit in the initial project, which gave me the confidence to pursue this idea.

used to kill several human beings when in the hands of a psychopath, it is used to prepare gourmet cuisine in the hands of a chef. Consequently, it would be an overreaction to stop selling this kitchen tool because it has been used as a murder weapon.

For similar reasons, it would be perhaps disproportionate to stop testing autonomous vehicles because a car, whether autonomous or human-driven, is an instrument at the service of a human. In other words, to be at the service of a human is to be at the service of an end. From a strictly economic point of view, the autonomous vehicle is a material good or means of transport; like any economic good or means, it is not produced for itself but to be bought and used by consumers. Therefore, the proper purpose of material goods in general and of the autonomous vehicle in particular is to be at the service of the potential buyer. This is the reason why the customer is king.

Being at the service of the customer brings to light the problem of the purpose of the company, which has been the subject of many theoretical controversies between on the one hand, the shareholder value model (Friedman, 1970 ; Hasnas, 1998 ; Stout, 2012; Hart & Zingales 2017; Robson, 2019) and the normative stakeholder theory (Freeman 1984; Freeman & Reed 1983, Freeman 1994 ; Freeman et al. 2010; Freeman, Harrison, & Zyglidopoulos 2018 ; Jones, Wicks, & Freeman 2002 ; Phillips, Freeman, & Wicks 2003), on the other hand. First, according to the value model, the firm has real moral obligations only to its shareholders. Secondly, for theories that give moral weight to stakeholders, the firm has a moral obligation to secure the interests of a broad range of communities, which shareholders are only one group. This article contributes to the debate between value theorists and stakeholder theorists, insisting on the purpose of the firm to serve all stakeholders of which the final consumer is one of the essential parts. Indeed, the corporate pursues the satisfaction of its customers, by the means that are the products or services. Therefore, a means is only valid if it leads to an end.

Thus, the term moral is defined by Aristotle in *Nicomachean Ethics* as the respect of the end. The word ‘**morals**’ is formed from the Latin *mos* in the plural *mores* and the Greek *ethos* from which is formed the term ‘**ethics**’ both describing human action. Furthermore, what allows humans to act freely is our intelligence. Therefore, the moral freedom or capacity to choose is found in humans only because we have intelligence.

We can apply the word ‘**morals**’ to material things legitimately provided we keep in mind the analogy of attribution that finds it. In reality, morality is attributed to the human act and so in a derived sense to material things. In itself, material things acquire morality only through how they are used by human beings, as I have already discussed above concerning the kitchen knife.

However all empirical papers that analyze the moral dilemmas of the autonomous vehicle in an accident situation assume that the autonomous vehicle is a moral agent in the strict meaning, i.e., it can choose who can live or die (Awad et al. 2018; Frank et al. 2019; Bigman and Gray, 2020; Awad et al. 2020; Bonnefon et al. 2016). Statistical tests are performed on the questionnaires based on the trolley problem. Initially, the trolley problem imagines a trolley about to run over five people, and the only way to spare their lives is to direct the trolley to another lane, where there is one person who will not have time to avoid being hit. Adapted in Moral Machine, the dilemma amounts to giving an autonomous vehicle the possibility to choose its victims in an accident situation: should it save a man or a woman? Should it sacrifice its passenger or the pedestrians? The question is thus posed between two horns, and no matter the solution, the conclusion is the same: the autonomous vehicle must kill. In other words, autonomous vehicles should be programmed to kill.

This conclusion from Moral Machine (2018) about the autonomous vehicle dilemma leads to a question: What kinds of things can be produced or sold? In other words, can an autonomous vehicle programmed to kill have a commercial outlet?

In general, results related to the marketing of autonomous vehicles are rare. Eggers' (2022) article discusses brands from the perspective of buying or leasing an autonomous car, but does not address the case of autonomous vehicles and the moral dilemmas associated with them more specifically. However unresolved moral dilemmas can be a deterrent to purchase. Indeed, no potential buyer will purchase an autonomous vehicle that is programmed to sacrifice him or her in case of an accident. This research is the first that speaks to the business impact of the moral dilemmas of autonomous vehicle in an accident situation by focusing on the purpose of any vehicle: to serve all market stakeholders.

Previous studies (Awad et al. 2018; Frank et al. 2019; Bigman and Gray, 2020; Awad et al. 2020; Bonnefon et al. 2016) only focus on solving the moral dilemmas of autonomous vehicles without considering the commercial ethics of autonomous vehicles.

The paper returns to this hypothesis shows that the autonomous vehicle is not a moral agent in the strict meaning, i.e., not being free to act, it cannot decide who should live or die. The decision is established by the algorithm created by humans. Thus, the methodology used by the empirical papers that rely on the trolley problem is not appropriate. Indeed, the methodology does not take into account the definition or the nature of the autonomous vehicle, which must be a means or a product manufactured by economy. Knowledge of the nature of the autonomous vehicle makes it possible to establish its programming, namely, to be a means of transport.

My approach to this paper is empirical, that is, based on observation. My research returns to the trolley problem adapted to the autonomous vehicle vehicle, which is the center of all empirical studies dealing with the moral dilemmas of autonomous vehicles. Section 2 considers how autonomous vehicles must be programmed: it is necessary to know first what an autonomous vehicle is or what its nature is. There is an order. It is necessary first to know what a thing is to know how it acts. Section 3 applies the definition of autonomous vehicles to the

various moral dilemmas on which the Moral Machine paper is based. Section 4 gives the implications for all stakeholders in the autonomous vehicle market.

Section 5 concludes how autonomous vehicles should be programmed, i.e., to serve and not to kill.

2 Nature and ethical's production of the autonomous vehicle in accident situation

Every being has a proper operation. Every operation or movement has a term that defines the nature of the being. The purpose of a thing corresponds to its nature; it corresponds to what the thing is. Thus, to program an autonomous vehicle (AV), it is necessary first to know what an autonomous vehicle is, i.e., what is its nature. Once we know the nature of the AV, we will be able to see how to program it: things act as they are. Thus there is an order, one must know what a thing is to know how it acts. Thus the answer of the action is always found in the nature of the thing.

A car with or without a driver is a means to an end: to transport things or humans from a point *A* (beginning) to a point *B* (end). Consequently, the natural purpose of the car with or without a driver is not to kill but to transport.

Before carrying out the act of transporting, AV must be produced by the corporate. The natural end of production is the potential customer for whom the act of production is ordered. Production is not carried out for its own sake but for the end buyer. Thus, the proper and immediate end of the thing produced is the customer. The economic good is thus produced for the customer. This is another way of saying that the customer is first in the economic order. Thus, the ethic of production or the ultimate goal of production is the final consumer. The

customer is what all the corporate's actions are aimed at: serving or transporting in the case of autonomous vehicles. As a result, production and consumption are not servants of each other in the same way. Consumption remains the end of production: we produce to consume more than we consume to produce. It is the customer who orders the production and not the other way around. The value proposed by a corporate is the response it personally provides to an identified need.

In contrast to value theorists who place the unique and ultimate purpose of the corporate in profit (Friedman, 1970 ; Hasnas, 1998 ; Stout 2012 ; Hart & Zingales 2017; Robson 2019), in the order of purposes, profit is not the ultimate purpose, but the intermediate or instrumental end. Indeed, as the term indicates, profit is the benefit provided to the corporate by its value proposition to the customer. Thus without a value proposition to the customer, there is no sustainable profit. Thus the profit or the financial returns of the corporate is the compensation of its service. That's why, as I said, the customer is king. It is to meet the needs and requirements of the king that corporates compete with each other. It is the meaning of the word compete, the corporates compete to satisfy the king. Thus, the object of the competition is the satisfaction of the customer.

Therefore, buyers must decide what they want to produce. Consequently, the production of AV, like any human operation, has a purpose: to transport its owner from point *A* to point *B*.

This principle of finality is explained by Aristotle in Book II of *Physics*. Any operation or movement, even a purely physical operation, is directed to an end. The object of the AV has a wide meaning and a strict meaning. In the wide meaning, the object includes all that the act concerns. In this sense, the object according to Aristotle in *Organon* includes the **circumstances or accidents** of the human act or in the case of the Moral Machine of the AV:

- **Who** : killing or saving a man; a child; a baby ; a woman; a pedestrian; a rich person; a poor person; etc;

- **Quantity:** kill one person to save five people; run over 30 people to save one passenger; save more people rather than fewer;
- **By what means:** autonomous vehicle;
- **Duration of the action:** experiment integrating time (Franck et al. 2019);
- **How was the action taken:** deliberation, intuition (Franck et al. 2019).

Accidents are thus elements external to the very nature of the human act or the machine. They presuppose the human act or the machine already constituted in its nature.

Therefore, the object is distinguished from the circumstances and is defined as that which is reached in the first place and directly by the act. Thus, it is the first and immediate term to which the act of the AV is ordered by its nature: to transport and not to kill. Changing the object of the AV by the act of killing, as in most Moral Machine's dilemmas, does not answer the end of the AV, which is to transport.

3 Solving the moral dilemmas of the autonomous vehicle

After having defined what a car is, a means, and having seen that its finality must be to transport and not to kill, it is now possible to confront the definition of AV with the different moral dilemmas in Moral Machine. The question is therefore between two horns or two terms, but the horns are always attached to the same beast, such is the definition of the dilemma according to Aristotle in *Organon*. Thus, regardless of the solution, the conclusion is the same, and the autonomous vehicle must kill :

1. **Sacrifice** passengers *or sacrifice* pedestrians;
2. **Sacrifice** men *or sacrifice* women;
3. **Sacrifice** the young *or sacrifice* the old;
4. **Sacrificing** humans *or sacrificing* animals;
5. **Sacrificing** "fit" people *or sacrificing* "unfit" people;
6. **Sacrificing** more lives *or sacrificing* fewer lives;
7. **Sacrificing** people according to their "higher social status" *or sacrificing* people of lower social status;
8. **Sacrifice** pedestrians crossing legally *or sacrifice* jaywalkers.

However, being programmed to kill does not achieve the natural purpose of the machine, which is to serve an end: to transport one or more people from point *A* to point *B*.

The examination of the last proposition in Moral Machine: *to go straight without changing direction*; affords some observations. First, this preference was measured as the weakest and therefore not statistically significant. Second, even if this last proposition was included by the authors as being part of the dilemma, it is not part of the dilemma according to the definition of the dilemma itself because the outcome is not certain. Finally, this proposal, *to go straight rather than change direction*, means in the paper Moral Machine, **not to act**. This preference is consistent as we already pointed out in the introduction, the autonomous vehicle is not a moral agent in the strict meaning, i.e., not being free to act, it cannot decide or choose who should live or die. Indeed, the decision is established by the algorithm created by humans. The only prudent action in an accident situation that respects the nature of the autonomous vehicle, i.e., being a means of transportation, is to maintain its trajectory, i.e., not to choose who should die or live.

This preference is also consistent with the moral psychology literature that has identified it as an ‘omission bias’ (Spranca et al, 1991; Cushman et al, 2006; Descloli et al, 2011). In other words, in moral choices, with equal consequences, people prefer not to act.

4 Implications for Stakeholders

This solution could be of interest to all major stakeholders of the future autonomous vehicle market:

- States, guarantor of the common good or of the social order which is here the road safety could apply the current legal rules applicable to cars with a human driver. Indeed, none of the current rules in any country give the vehicle the power to choose its victims in an accident situation; because this would lead to social disorder and therefore the common good which is road safety would no longer be achieved. However, it is the main mission of the State to ensure the realization of the social order or the common good: road safety;
- This solution could also be of interest to autonomous vehicle manufacturers whose cars do not have the freedom to choose one human life over another. They will have to continue to focus on safety for all road users. Road safety is the common good of the community, which is the main task of public authorities. Corporates that are run by intelligent and therefore free human beings, perform free acts of production that belong to them; as the profits generated by the activity also belong to the corporates. Therefore, they can be held responsible for their acts; because they make free acts. It is because corporates are free to produce, that their acts belong to them. This is the nominal

definition of the notion of responsibility applied to corporates. Because, as previously mentioned, it is humans who produce. The “Corporate social responsibility”, or, CSR is explained by the fact that the acts of production of the company have repercussions in the communities in which they live (Bowen, 1953). Therefore, the manufacturers of autonomous vehicles, as actors of the city, must continue to contribute to the common good or to the order of the city. Their contribution to the common good or to the order in the city, consists in producing objects and in this case autonomous vehicles which must not be a factor of disorder or danger for any of the components of the city. This is how the notion of corporate social responsibility applied to the manufacturers of autonomous vehicles comes about, or in other words, their contribution to the common good: road safety;

- Potential buyers of autonomous vehicles could buy a car that respects its purpose: to be a means.

5 Conclusion

No AV buyers would buy a car that is programmed to kill its driver because this is not the nature or the purpose of the autonomous vehicle. To be at the service of the buyer is to respect the end: to transport and not to kill its owner. In fact, in a competitive market , a technology prevails over competing technologies when it is favored by customers. This is the main lesson of this paper and the solution to moral dilemmas: respecting the objective, i.e. serving all stakeholders, of which the final consumer is one of the key parts.

Thus, the machine is not a moral agent in the strict meaning, because its actions are determined by the algorithm produced by human. In fact, since the autonomous vehicle is not free, it cannot choose who should live or die. Indeed, even at Tesla, a pioneer in the sector with its famous Autopilot, the driver must always keep his or her hands on the wheel, as he or she must be able to regain control at any time, or even correct a potential system error.

Therefore, to know how to program an autonomous vehicle, one must first know what a car with or without a driver is, i.e., a means to an end. It is clear that programming to kill cannot be the purpose of an autonomous vehicle. The economic purpose of the vehicle corresponds to the morality (Sen, 1987) of being at the service of a purpose: to transport one or more people from point *A* to point *B*.

References

Althausen, J. (2017). Moral programming will define the future of autonomous transportation.

VentureBeat, <https://venturebeat.com/2017/09/24/moral-programming-will-define-the-future-of-autonomous-transportation/> (2017)

Awad, R. *et al.* (2018). The Moral Machine experiment. *Nature*. <https://doi.org/10.1038/s41586-018-0637-6>.

Awad, E. *et al.* (2020). Reply to: Life and death decisions of autonomous vehicles. *Nature*. <https://doi.org/10.1038/s41586-020-1987-4>.

Bigman, Y-E. & Gray, K. (2020). Life and death decisions of autonomous vehicles. *Nature*. <https://doi.org/10.1038/s41586-018-0637-6>.

Bonnefon, J-F *et al.* (2016). The social dilemma of autonomous vehicles. *Science* **352**, 1573-1576.

Bowen H.R. (1953), *Social Responsibilities of the Businessman*, Harpet & Row.

Cushman F., Young L. & Hauser, M. (2006) The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science* **17**, 1082-1089.

Descloli, P., Christner, J. & Kurzban, R. (2011)The omission strategy. *Psychological Science* **22**, 442-446.

Eggers, F. & Eggers, F. Drivers of autonomous vehicles—analyzing consumer preferences for self-driving car brand extensions (2022). *Marketing Letters*, 33 (8).

Foot P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*. **5**, 5-15.

Frank, D. A., Chrysochou, P., Mitkidis, P. & Ariely, D. (2019). Human decision-making biases in the moral dilemmas of autonomous vehicles. *Nature*.

<https://doi.org/10.1038/s41598-019-49411-7>.

Freeman, R.E. & D.L. Reed. (1983). Stockholder and Stakeholders: A New Perspective on Corporate Governance. *California Management Review*, 25(3): 88–106.

Freeman, R.E. (1984). *Strategic Management: A Stakeholder Approach*, Boston, MA: Pitman.

Freeman, R.E. (1994). The Politics of Stakeholder Theory: Some Future Directions. *Business Ethics Quarterly*, 4(4): 409–421.

Freeman, R.E., J.S. Harrison, A.C. Wicks, B.L. Parmar, & S. De Colle. (2010). *Stakeholder Theory: The State of the Art*, Cambridge: Cambridge University Press.

Freeman, R.E., J.S. Harrison, & S. Zyglidopoulos. (2018). *Stakeholder Theory: Concepts and Strategies*, New York: Cambridge University Press.

- Friedman, M. (1970). The Social Responsibility of Business is to Increase its Profits. *New York Times Magazine* (September 13): 32–33, 122–124.
- Hasnas, J. (1998). The Normative Theories of Business Ethics: A Guide for the Perplexed. *Business Ethics Quarterly*, 8(1): 19–42.
- Jones, T. M., A.C. Wicks, & R.E. Freeman. (2002). Stakeholder Theory: The State of the Art, in N.E. Bowie (ed.), *The Blackwell Guide to Business Ethics*, Malden, MA: Blackwell, pp.19–37
- Martin, K., Shilton, K, & J. Smith. (2019). Business and the Ethical Implications of Technology: Introduction to the Symposium. *Journal of Business Ethics*, 160.
- Marcoux, A. (2009). Business-Focused Business Ethics. in *Normative Theory and Business Ethics*. J. Smith. Plymouth Rowman & Littlefield: pp. 17–34 [ISBN 0-7425-4841-4](#)
- McCausland, P. Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk. *NBC News* (2019).
- Meder, B., Fleischhut, N., Krumnau, N-C & Waldmann, M.R. (2019). How Should Autonomous Cars Drive? A Preference for Defaults in Moral Judgments Under Risk and Uncertainty. *Risk Analysis* **39**, 295-314.
- Robson, G. (2019). To Profit Maximize, or Not to Profit Maximize: For Firms, This is a Valid Question. *Economics and Philosophy*, 35(2): 307–320.
- Sen, A. *On ethics and Economics*, Oxford, Wiley-Blackwell (1987).
- Spranca, M., Mink, E. & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology* **1**, 76-105.
- Stout, L.A. (2012). *The Shareholder Value Myth: How Putting Shareholders First Harms Investors, Corporations, and the Public*, San Francisco, CA: Berrett-Koehler Publishers.
- Zingales, L., & O. Hart. (2017). Companies Should Maximize Shareholder Welfare not Market Value. *Journal of Law, Finance, and Accounting*, 2(2): 247–275.

