



AK2003 TECHNOLOGY AND ETHICS

Who is morally responsible for fully autonomous cars?

Martin Barksten
barksten@kth.se

20th May 2015

1 Introduction

En sak som diskuteras och arbetas mycket med just nu är så självkörande bilar, alltså bilar som styrs av ett autonomt system istället för en människa. Ett autonomt system är ett system som kan utföra en uppgift utan mänsklig hjälp. Flertalet företag håller just nu på att utveckla bilar som är delvis eller fullt autonoma. Mest intressant ur ett filosofiskt perspektiv är de bilar som NHTSA definierar som nivå 4: fullständigt autonoma bilar som inte kräver någon mänsklig hjälp [1]. Flera företag arbetar med autonoma bilar på nivå 4, och exempelvis Googles autonoma bilar har kört över 1 miljon kilometer de senaste åren [8]. Sergey Brin, en av Googles grundare, har i en intervju år 2012 sagt "You'll ride in robot cars within 5 years" [5].

Autonoma system kan minimera mänskliga fel och därmed minska riskerna för människor men introducerar andra, främst moraliska, problem. Systemen tillåter oss att bestämma vad ett korrekt agerande givet en viss situation är, på gott och ont. I fallet med autonoma bilar är det ofta positivt: exempelvis kan vi bestämma att bilen ska bromsa vid gult ljus, inte accelerera för att hinna innan det blir rött. Det blir dock svårare när det krävs att systemen tar beslut i moraliskt komplexa situationer.

Ett exempel på ett problem en autonom bil kan ställas inför är en variant av det kända "trolley problem" [21]; Noah J. Goodall presenterar en situation där bilen kör på en trång bro och i motsatt riktning kommer en buss, som plötsligt svänger över mot bilens sida av vägen [6]. Bilen ställs då inför tre möjliga alternativ:

1. Svänga av bron och garantera en allvarlig olycka med enbart bilen;
2. Frontalkrocka med bussen och orsaka en mindre allvarlig olycka involverande både bil och buss;
3. Eller försöka köra förbi bussen till höger om den med en möjlighet att undvika en krock, men med en risk för en värre olycka med större skaderisk för passagerarna i både bilen och bussen.

Situationen är en form av Kobayashi Maru test [19], ett scenario där alla möjliga alternativ är dåliga på olika sätt. En människa som ställs inför problemet följer troligen sin egna och individuella moral för att välja hur den agerar. För ett autonomt system krävs dock att vi på förhand programmerar systemet och bestämmer hur det ska agera, vilket kräver att vi ger systemet någon form av moral.

Autonoma bilar kommer ställas inför verkliga varianter av Kobayashi Maru testet. Noah J. Goodall visar i en artikel att bilarna kommer, på grund av egna fel eller yttre omständigheter som inte går att kontrollera, hamna i olyckor [7]. Med tanke på det stora antalet bilar – i exempelvis USA hade 190 miljoner personer körkort år 2000 [3] och körde i snitt 21,7 tusen kilometer per år [2] – kommer dessa scenarion troligen inte vara ovanliga.

2 Essay question

Autonoma bilar har flera för- och nackdelar rent praktiskt, men framförallt av intresse är en moralisk diskussion kring bilarna. Min frågeställning är därför:

Är det moraliskt att tillåta autonoma bilar?

3 Discussion

De moraliska problemen med autonoma bilar har diskuterats av flera olika författare. Problematiken kring att implementera en moral i autonoma system tar Duncan Purves, Ryan Jenkins och Bradley J. Strawser upp i en artikel med fokus på autonoma vapensystem [15]. Don Howard skriver i en bloggpost om de tre stora fördelar han ser med autonoma bilar och kommer till slutsatsen att det är omoraliskt att inte så snabbt som möjligt introducera dem på marknaden [11]. Patrick Lin diskuterar å andra sidan de moraliska problem som kan finnas med självkörande bilar [12]. Även problematiken kring vem som står ansvarig när krockar sker har diskuterats av flera olika personer från olika perspektiv, både filosofiska [6, 7, 10] och rättsliga [13].

4 Conclusion

Med hjälp av Sven Ove Hanssons modell kan vi ge ett svar på om det är moraliskt riktigt att tillåta självkörande bilar: om vi kan argumentera för att de positiva konsekvenserna för var och en av de utsatta individerna är tillräckliga för att väga upp för de kränkningar av rättigheter ett tillåtande av autonoma bilar innebär. Vi måste alltså kunna motivera att de individer som behandlas omoraliskt av autonoma bilar trots allt får ut mer av de positiva konsekvenserna av att bilarna räddar liv och minskar skador.

References

- [1] National Highway Traffic Safety Administration. “Preliminary Statement of Policy Concerning Automated Vehicles”. English. In: (2013). URL: http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf.
- [2] Federal Highway Administration. *Average Annual Miles per Driver by Age Group*. 2015. URL: <https://www.fhwa.dot.gov/ohim/onh00/bar8.htm> (visited on 01/07/2016).
- [3] Federal Highway Administration. *Licensed Drivers*. 2015. URL: <http://www.fhwa.dot.gov/ohim/onh00/onh2p4.htm> (visited on 01/07/2016).
- [4] J.-F. Bonnefon, A. Shariff, and I. Rahwan. “Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?” In: *ArXiv e-prints* (Oct. 2015). arXiv: 1510.03346 [cs.CY].

- [5] CNET. *Google's Sergey Brin: You'll ride in robot cars within 5 years*. 2012. URL: <http://www.cnet.com/news/googles-sergey-brin-youll-ride-in-robot-cars-within-5-years/> (visited on 12/03/2015).
- [6] Noah J Goodall. "Ethical Decision Making During Automated Vehicle Crashes". In: *Transportation Research Record: Journal of the Transportation Research Board* 2424 (2014), pp. 58–65.
- [7] Noah J. Goodall. "Machine Ethics and Automated Vehicles". English. In: *Road Vehicle Automation*. Ed. by Gereon Meyer and Sven Beiker. Lecture Notes in Mobility. Springer International Publishing, 2014, pp. 93–102. ISBN: 978-3-319-05989-1. DOI: 10.1007/978-3-319-05990-7_9. URL: http://dx.doi.org/10.1007/978-3-319-05990-7_9.
- [8] Google. *Google Self-Driving Cars Project*. 2015. URL: <https://www.google.com/selfdrivingcar/> (visited on 12/03/2015).
- [9] Sven Ove Hansson. "Ethical Criteria of Risk Acceptance". English. In: *Erkenntnis* 59.3 (2003), pp. 291–309. ISSN: 0165-0106. DOI: 10.1023/A:1026005915919.
- [10] Alexander Hevelke and Julian Nida-Rümelin. "Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis". English. In: *Science and Engineering Ethics* 21.3 (2015), pp. 619–630. ISSN: 1353-3452. DOI: 10.1007/s11948-014-9565-5. URL: <http://dx.doi.org/10.1007/s11948-014-9565-5>.
- [11] Don Howard. *Robots on the Road: The Moral Imperative of the Driverless Car*. 2013. URL: <http://donhoward-blog.nd.edu/2013/11/07/robots-on-the-road-the-moral-imperative-of-the-driverless-car/#.Vo1LnZMrLv2> (visited on 01/06/2016).
- [12] Patrick Lin. *The Ethics of Saving Lives With Autonomous Cars Is Far Murkier Than You Think*. 2013. URL: <http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars/> (visited on 01/06/2016).
- [13] Gary E. Marchant and Rachel A. Lindor. "The Coming Collision Between Autonomous Vehicles and the Liability System". English. In: *Santa Clara Law Review* 52.4 (2012), pp. 1321–1340. URL: <http://digitalcommons.law.scu.edu/lawreview/vol52/iss4/6>.
- [14] World Health Organization. "Global Status Report on Road Safety 2013: Supporting a Decade of Action". English. In: (2013). URL: http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/.
- [15] Duncan Purves, Ryan Jenkins, and Bradley J. Strawser. "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons". English. In: *Ethical Theory and Moral Practice* 18.4 (2015), pp. 851–872. ISSN: 1386-2820. DOI: 10.1007/s10677-015-9563-y. URL: <http://dx.doi.org/10.1007/s10677-015-9563-y>.

- [16] Brandon Schoettle and Sivak Michael. “A preliminary analysis of real-world crashes involving self-driving vehicles”. English. In: (2015). URL: <http://www.umich.edu/~umtriswt/PDF/UMTRI-2015-34.pdf>.
- [17] Stanford Center for the Study of Language and Information. *Deontological Ethics*. 2012. URL: <http://plato.stanford.edu/entries/ethics-deontological/> (visited on 01/07/2016).
- [18] Stanford Center for the Study of Language and Information. *Deontological Ethics*. 2012. URL: <http://plato.stanford.edu/entries/consequentialism/> (visited on 01/07/2016).
- [19] Wikipedia. *Kobayashi Maru*. 2016. URL: https://en.wikipedia.org/wiki/Kobayashi_Maru (visited on 01/07/2016).
- [20] Wikipedia. *Three Laws of Robotics*. 2016. URL: https://en.wikipedia.org/wiki/Three_Laws_of_Robotics (visited on 01/07/2016).
- [21] Wikipedia. *Trolley Problem*. 2015. URL: https://en.wikipedia.org/wiki/Trolley_problem#cite_note-Philippa_Foot_1978-1 (visited on 12/03/2015).