From Trolleys to Risk: Models for Ethical Autonomous Driving

Noah J. Goodall

Pre-print version. Published in *American Journal of Public Health,* 107(4), April 2017, pp. 496.
Available at: http://dx.doi.org/10.2105/AJPH.2017.303672

The article by Fleetwood[1] in this issue of *AJPH* provides an overview of the public health implications of highly-automated vehicles, with a focus on the ethics of a vehicle's behavior when a crash is unavoidable, i.e. its "ethical crashing algorithms." While autonomous vehicles are widely-expected to reduce crash rates, those benefits may not be distributed equitably, and some users may receive more benefit than others. Just as airbags save many, they also kill a few that would otherwise not have died. This creates a smaller but persistent public health issue, and the authors provide a helpful exploration of the unique ethical challenges created by the (hopefully) rare autonomous vehicle crashes.

Assigning values

As the authors point out, decisions about how to respond in crash situations require that the vehicle, and by proxy its designers and software developers, assign values to various objects – otherwise, the car would treat a traffic cone and a pedestrian identically. At a broader level, I would argue that *all* driving, not just pre-crash driving, requires assigning values to different objects. How much space to give a cyclist as it passes, how much it slows down in a residential neighborhood, these decisions require the vehicle to balance the safety of its own passengers and road users, and to balance safety and time. These subtler decisions will affect safety. I expect truck-AV crashes to decrease faster than other crash types, as automated vehicles will position themselves laterally in a lane farther from larger trucks and therefore closer to smaller cars[2]. While the authors focused on pre-crash maneuvers, there is value in researching and discussing the role of ethics in general driving.

Trolley problem

I caution against an over-reliance on the trolley problem as representative of the ethics of autonomous vehicle behavior. The trolley problem is a thought experiment borrowed from philosophy where a trolley is on a collision course with five people, but can be diverted onto another track where it will kill only one. It has many variations, and has been used to explore ethical principles that may underlay people's intuitions about the most moral course of action. Similar scenarios for autonomous vehicles have been proposed, where vehicles must choose between killing two people in the street and one on the sidewalk, a pedestrian and the autonomous vehicle's passenger, or a child and some large number of dogs. The trolley problem is useful in discussions because it is fairly well-known, it represents clear choice with only two distinct alternatives, and assumes completely certain outcomes with obvious moral consequences[3]. These attributes strike vehicle developers as unrealistic and naive[4]; real driving dilemmas have many subtle choices, uncertain outcomes, and often an obviously superior course of action, e.g. apply the brakes.

Merging ethical systems

The ethics of automated vehicles may be better framed as the fair distribution of risks, both during *and prior to* forced-choice situations. The distribution of risks, or the rationing of benefits, has been addressed in previous public health challenges, including vaccine rationing[5], organ donation[5], radiation exposure[6], and even the military draft[5]. In each scenario, officials have used approaches that integrate elements of utilitarian and deontological ethics to create a justifiable system. For example, many organ donation programs combine deontological ethics (first-come-first-served) with utilitarian ethics (sickest-first) when prioritizing recipients. Merging ethical systems may be unpopular with philosophers, but it seems to work well in practice.

Open dialogue

The common thread among most successful risk management approaches in public health seems to be an open discussion, or at least some transparency, regarding the reasoning behind the chosen system. Even better is to include widely-accepted ethical principles, and ideally the language of ethics, when presenting crash avoidance and general driving safety algorithms. While I find the authors' suggestion of additional state and federal regulation on this topic premature, an open dialogue about safety is certainly appropriate.

References

1. Fleetwood J. Public health, ethics, and autonomous vehicles. *Am J Public Health.* 2017;107(4):532-537. doi:10.2105/AJPH.2017.303672
2. Goodall NJ. Can you program ethics into a self-driving car? *IEEE Spectrum.* 2016;53(6):28-58. doi:10.1109/MSPEC.2016.7473149.
3. Goodall NJ. Away from trolley problems and toward risk management. *Applied Artificial Intelligence.* 2016;30(8):810-821. doi:10.1080/08839514.2016.1229922.
4. Rose B. The Myth of Autonomous Vehicles' New Craze: Ethical Algorithms. *TechCrunch.* http://social.techcrunch.com/2015/11/23/the-myth-of-autonomous-vehicles-new-craze-ethical-algorithms/. Accessed February 12, 2016.
5. Persad G, Wertheimer A, Emanuel EJ. Principles for allocation of scarce medical interventions. *The Lancet.* 2009;373(9661):423-431. doi:10.1016/S0140-6736(09)60137-9.
6. Hansson SO. Ethics and radiation protection. *J Radiol Prot.* 2007;27(2):147. doi:10.1088/0952-4746/27/2/002.