Transportation Research Part C 80 (2017) 206–215

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

When human beings are like drunk robots: Driverless vehicles, ethics, and the future of transport

Robert Sparrow *, Mark Howard

Department of Philosophy, Monash University, Victoria 3800, Australia

ARTICLE INFO

Article history:
Received 27 December 2016
Received in revised form 23 April 2017
Accepted 24 April 2017

Keywords:
Autonomous cars
Driverless vehicles
Ethics
Urban planning
Safety
Transport policy

ABSTRACT

It is often argued that driverless vehicles will save lives. In this paper, we treat the ethical case for driverless vehicles seriously and show that it has radical implications for the future of transport. After briefly discussing the current state of driverless vehicle technology, we suggest that systems that rely upon human supervision are likely to be dangerous when used by ordinary people in real-world driving conditions and are unlikely to satisfy the desires of consumers. We then argue that the invention of fully autonomous vehicles that pose a lower risk to third parties than human drivers will establish a compelling case against the moral permissibility of manual driving. As long as driverless vehicles aren't safer than human drivers, it will be unethical to sell them. Once they are safer than human drivers when it comes to risks to 3rd parties, then it should be illegal to drive them: at that point human drivers will be the moral equivalent of drunk robots. We also describe two plausible mechanisms whereby this ethical argument may generate political pressure to have it reflected in legislation. Freeing people from the necessity of driving, though, will transform the relationship people have with their cars, which will in turn open up new possibilities for the transport uses of the automobile. The ethical challenge posed by driverless vehicles for transport policy is therefore to ensure that the most socially and environmentally beneficial of these possibilities is realised. We highlight several key policy choices that will determine how likely it is that this challenge will be met.

1. Introduction

Every presentation by an engineer on the topic of driverless vehicles that we have seen has begun with a statistic about road crash fatalities. Worldwide, 1.25 million people are killed annually by people driving cars (World Health Organization, 2015, p. 2). An oft-cited statistic is that over 90 percent of all road traffic accidents are a result of human error and behaviour (Anderson et al., 2016; Gao et al., 2014; National Highway Traffic Safety Administration, 2008). Driverless cars, we are solemnly informed, can do better. They will save lives (Anderson et al., 2016; Garza, 2011, p. 606). The case for driverless vehicles is ultimately, then, an ethical one. In this paper, we want to take this claim seriously and show that it has radical implications for the future of transport. Our argument proceeds via a simple dialectic. As long as driverless vehicles aren't safer than human drivers, it will be unethical to sell them (Shladover, 2016). Once they are safer than human drivers when it comes to risks to 3rd parties, then it should be illegal to drive them: at that point human drivers will be the moral equivalent of drunk robots. Freeing people from the necessity of driving, though, will transform the relationship people have with

* Corresponding author.
E-mail address: robert.sparrow@monash.edu (R. Sparrow).

http://dx.doi.org/10.1016/j.trc.2017.04.014

their cars, which will in turn open up new possibilities for the transport uses of the automobile. The challenge posed by driverless vehicles for transport policy is therefore to ensure that the most socially and environmentally beneficial of these possibilities is realised.

In the next section of the paper, we briefly discuss the current state of driverless vehicle technology and argue that systems that rely upon human supervision are likely to be dangerous when used by ordinary people in real-world driving conditions and are unlikely to satisfy the desires of consumers.[1] For this reason, we anticipate that the future of driverless vehicles will be vehicles that are fully autonomous and do not require human supervision.[2] In Section 3, we argue that the invention of fully autonomous vehicles that pose a lower risk to third parties than human drivers will establish a compelling case against the moral permissibility of manual driving. We also highlight two plausible mechanisms whereby this ethical argument may generate political pressure to have it reflected in legislation. In Section 4, we acknowledge some complexities that Section 3 neglects for the sake of ease of exposition. In the fifth section, we offer some predictions about the implications of the adoption of autonomous vehicles for the future of the automobile and of the city. In the sixth and final section, we highlight several key policy choices that will determine whether or not the best outcomes made possible by the development of autonomous vehicles will be realised.

2. Safer at any speed: the argument for driverless vehicles

Driverless vehicles are widely anticipated to represent the future of transport (Bamonte, 2013; Bilger, 2013; Burns, 2013). The date at which we may expect their arrival, however, is the subject of some dispute. According to some authorities, driverless cars are possible already and will be on the market by 2020 (Fagnant and Kockelman, 2014, p. 1; Burns, 2013, p. 182). Other writers are more cautious and suggest that it will be perhaps another two or three decades before the technology of driverless vehicles is mature enough to be suitable for widespread use (Litman, 2015, p. 1; Shladover, 2016). The explanation for this difference in expectations regarding the timeframe for the introduction of driverless vehicles lies in a difference in opinion regarding the extent of the technological challenges that will need to be overcome in order for them to safely integrate into the transport environment.

One popular theory about how driverless vehicles will take their place on the roads involves gradually increasing levels of automation of tasks currently performed by humans while still retaining human supervision of the driving task. According to this way of thinking, cruise control and anti-skid braking systems, already widely adopted, represent the beginnings of autonomous driving (Garza, 2011, p. 584; Fagnant and Kockelman, 2015a, p. 168; Gordon and Lidberg, 2015, p. 959); lane change assist and automated freeway driving represent the next step, with Tesla's "autopilot" most of the way to realising this latter goal (Singhvi and Russell, 2016; Vlasiv, 2016). By progressively extending the capacities of these technologies, engineers will be able to produce vehicles that can drive in a larger and larger set of environmental and road conditions, as long as a human being is available to step in should conditions exceed the capacities of the vehicle to handle them safely (Gordon and Lidberg, 2015).[3]

The driverless vehicles that are being tested on the roads today almost all rely upon supervision of a human being in order to drive at speed under real-world conditions in urban environments (Thrun, 2010; Goodall, 2014a, p. 58; Google, 2016). That is to say, they have only Level 3 or Level 4 autonomy. However, there is an obvious problem with this approach, which we believe will be extremely difficult to overcome: at some point in the not-too-distant future, when driver assist systems become sufficiently reliable, the human "supervisor" will stop paying attention. Human beings quickly cease to pay attention to – or even to perceive – features of their environment that are not relevant to the tasks in which they are engaged (Merat and Jamson, 2009). To a certain extent this dynamic may be resisted by trained professionals, such as airline pilots, or by engineers who are being paid to monitor the activities of driverless vehicles in order to improve them. Nevertheless, an ordinary person "supervising" the activities of an autonomous vehicle that is 99.9% reliable will quickly cease to be engaged in this task. If the vehicle then requires them to take control quickly when circumstances exceed the capacity of the driving software to manage them safely, the human supervisor is unlikely to be able to do so effectively (Shladover, 2016). It may take more than 30s – and will at minimum take 2s – for a human being to regain situational awareness when required to do so (Lin, 2015, p. 72; Kirkpatrick, 2015, p. 20; Radlmayr et al., 2014).

It might be thought that as long as the autonomous driving systems fail gradually and gracefully it will be possible to arrange for a human driver to retake control of the vehicle as required.[4] Driverless vehicles need not be able to cope with all road conditions, for instance, in order to be useful. For example, a car that is capable of driving on freeways but not local

[1] The SAE international standard for defining levels of automation, which we adopt in this article, ranges from no automation (level 0) to full automation (level 5). Levels 1–3 require human supervision and human takeover when the driving task is beyond the capacity of the automated vehicle system, while levels 4–5 operate without human assistance (level 4 in certain environments). See: (NHTSA, 2016; Shladover, 2016).

[2] Unless otherwise noted, when discussing fully autonomous vehicles we are referring to vehicles with SAE level 4 automation, where "an automated system can conduct the driving task and monitor the driving environment, and the human need not take back control ..." (NHTSA, 2016, p. 9), capable of driving in North American and/or European cities and on sealed country roads. While also relevant, achieving SAE level 5 automation where "the automated system can perform all driving tasks, under all conditions that a human driver could perform them," (NHTSA, 2016, p. 9) is unnecessary for the success of our argument. By way of explanation, an SAE level 4 automated vehicle may be capable of performing all driving tasks on all designated roads, while being incapable of off-road driving, with the latter preventing its designation as SAE level 5.

[3] This appears to be the pathway to the introduction of driverless vehicle technology being pursued by Tesla. See: (Bhuiyan, 2016).

[4] For a discussion of 'safe' handover scenarios, see: (Gold et al., 2013).