Safety Science 102 (2018) 68–78

Contents lists available at ScienceDirect

Safety Science

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

Review

Automated driving: Safety blind spots

Ian Y. Noy[a,*], David Shinar[b], William J. Horrey[a]

[a] Liberty Mutual Institute for Safety, Hopkinton, MA, USA
[b] Ben Gurion University of the Negev, Be'er Sheva, Israel

ARTICLE INFO

Keywords:
Automated driving
Safety
Driver-vehicle interaction
Psychology
Autonomous vehicles

ABSTRACT

Driver assist technologies have reached the tipping point and are poised to take control of most, if not all, aspects of the driving task. Proponents of automated driving (AD) are enthusiastic about its promise to transform mobility and realize impressive societal benefits. This paper is an attempt to carefully examine the potential of AD to realize safety benefits, to challenge widely-held assumptions and to delve more deeply into the barriers that are hitherto largely overlooked. As automated vehicle (AV) technologies advance and emerge within a ubiquitous cyber-physical world they raise additional issues that have not yet been adequately defined, let alone researched. Issues around automation, sociotechnical complexity and systems resilience are well known in the context of aviation and space. There are important lessons that could be drawn from these applications to help inform the development of automated driving. This paper argues that for the foreseeable future, regardless of the level of automation, a driver will continue to have a role. It seems clear that the benefits of automated driving, safety and otherwise, will accrue only if these technologies are designed in accordance with sound cybernetics principles, promote effective human-systems integration and gain the trust by operators and the public.

1. Introduction

1.1. Emerging trends

Over the past 35 years, a great deal of scientific, engineering and media attention has been directed towards the rapidly evolving area of automated driving (AD), with increasing interest in the recent introduction of driverless vehicles. We define AD as driving in which at least some aspects of the dynamic driving tasks occur without driver input. AD has evolved rapidly due to advances in microprocessors, sensors, geodetic information systems, telecommunications and related technologies. Indeed, the rapidity of technological innovations in AD since the advent of Intelligent Transport Systems (ITS) in the 1980 s (Noy, 1989; Noy, 1997) has inspired a variety of disruptive paradigms for personal mobility.

How AD will continue to evolve has been the subject of much speculation among experts (VDA, 2015) but it seems clear that the exact evolution is interrelated with other important changes now taking place. For example, companies that have primarily specialized in digital and web-based technologies, such as Google, are entering a market that has traditionally been dominated by the automotive industry. This trend is introducing novel approaches to the design, usability and utility of motor vehicles. Moreover, there is little doubt that the smart car of the future will be an extension of the ubiquitous digital world, rather than the relatively independent mechanical platform that it is today. Finally, there is a growing trend towards deriving transportation-related services via the shared economy that challenges the hitherto dependence on car ownership for mobility.

There are generally two classes of on-board technologies. The first, advanced driver assistance systems (ADAS), is a broad category of devices that are intended to assist the driver in performing the driving task such as intelligent cruise control or electronic stability control. The second class, infotainment systems, are devices that enable the driver to perform auxiliary tasks such as voice communication, or interact with travel-related services and a variety of web-based apps. The scope of this paper is limited to ADAS, which, in theory, range in level of automation from single function technologies such as automatic braking to integrated system functions that support completely AD.

1.2. Penetration of AD into the fleet

Recently, the Society of Automotive Engineers (SAE) published an updated standard that defines five levels of AVs, ranging from no automation (driver in full control) to full automation (vehicle in full control at all times). The intermediate levels are distinguished by the number of automated systems and the required availability of the driver. Further details can be found in the SAE Automated Driving Standard (SAE, 2016) and the NHTSA Automated Vehicles Policy

* Corresponding author at: Liberty Mutual Research Institute for Safety, 71 Frankland Road, Hopkinton, MA 01748, USA.
E-mail address: iannoy@outlook.com (I.Y. Noy).

http://dx.doi.org/10.1016/j.ssci.2017.07.018
Received 4 March 2017; Received in revised form 5 July 2017; Accepted 31 July 2017
0925-7535/

(NHTSA, 2016).

While the SAE classification scheme has some utility—for example, it provides a basis for common terminologies across domains and functions (e.g., policymakers, engineers, researchers)—it does not adequately address aspects of driver-vehicle interactions from a human-centric perspective (cf. Parasuraman et al., 2000). More pointedly, it does not provide clear guidance in (a) establishing design requirements and helping designers make design decisions or tradeoffs, (b) helping regulators and the driving public to adequately understand system capabilities and limitations, and (c) informing industry and consumers about training needs.

While the accelerated development and implementation of ADAS systems is remarkable, there are widely divergent views concerning the timeline for the introduction of fully AD, ranging from one year to 40 years or longer. The IEEE (2014) predicts that AVs will account for up to 75% of vehicles on the road by the year 2040, with much of the time needed to socialize the technology (i.e., introduce the underlying technological concepts to the public, and reach widespread awareness of AD capabilities and limitations) and gain user acceptance. Indeed, fully automated vehicles already exist under limited use conditions (e.g., confined to areas such as airports in which special vehicles operate without a driver present to provide transportation services to terminals or parking lots). Of course, even if vehicles that can be operated completely in driverless mode anytime and everywhere were to become technologically feasible in the very near future, it is not certain to what extent they would achieve user acceptance or uptake or whether or when they might be offered or mandated on new vehicles. Market penetration rates under voluntary procurement are subject to a host of unknown social and economic factors. Unless they are mandated it is difficult to predict what portion of the fleet will be equipped with fully automated technologies. On the other hand, regulation of AVs is problematic. If regulators were to consider mandating AD, it would be a daunting task to develop the regulations and related test protocols and criteria as well as produce the requisite rulemaking justification based on extensive cost-benefit analyses. A further complication is that technology will continue to evolve, potentially rendering the test protocols under development obsolete before they can be implemented. If and when mandated, it would take many more years before AVs would constitute a significant proportion of the vehicles on the road due to vehicle fleet turnover rates. A scenario in which fully automated vehicles were mandated would disrupt the market and affect cost, performance and market penetration.

1.3. Driver-centric considerations

As will be discussed below, for the foreseeable future at least, the driver will continue to have a role in operating the vehicle (i.e., be in the loop) either as controller or supervisor or both. Of course, even in the context of fully AD, drivers may be considered as having a role if they have the option of turning off the automation and assuming manual control. This may be driven by Original Equipment Manufacturer (OEM) philosophy, consumer demand, or applicable regulations, which may vary regionally.

Curiously, there is little controversy over the predicted benefits of AD in terms of safety, environmental impact, fuel consumption and mobility. Some have identified barriers to achieving these benefits and formulated policy recommendations (Fagnant and Kockelman, 2015). With rare exception (Schoettle and Sivak, 2015), the projected benefits are accepted uncritically on the basis of industry claims. In this paper, we narrow the discussion to the issue of safety, and re-examine the benefits espoused by industry experts. In particular, we examine a variety of issues related to AD from a human-centric perspective. More specifically, we examine the role of the human driver, the potential ironies of automation, the need to take into account the diversity of the driving environments (including AVs with different levels of automation, dumb vehicles and other users), the need to consider AD in the context of sociotechnical systems theory, the emerging challenges associated with the economy of things, and the criticality of public trust in institutions as a precursor to gaining trust in AD.

2. Re-examining the safety benefits of AD

2.1. Eliminating driver error

Half a century ago, the common wisdom among Detroit automotive engineers was that it was possible to trace nearly all road crashes to a single component, or nut: the “nut” behind the wheel. The Indiana University “Tri-Level Study of Accident Causes” that was published in 1979 seemed to validate this when it concluded that human error was the probable or definite cause of over 90 percent of all crashes. Other studies that relied on a similar post-crash clinical analysis approach (Shinar, 2017), conducted in different parts of the world, reached similar conclusions (e.g., Hendricks et al., 2001; Otte et al., 2009; Sabey and Staughton, 1975; Singh, 2015). Even the most recent comprehensive naturalistic driving study conducted in the U.S., the SHRP2 study, using online recorders that documented nearly 1000 crashes, concluded that nearly 90% of the crashes were due to human error (Dingus et al., 2016).

Proponents of AD deduce that taking the human out of the loop should reduce the incidence of crashes by 90%. That is, replacing the driver with state-of-the-art sensing, computing, communication, and robotic-control technologies, would eliminate human error as a causative factor. However, this deduction warrants a more detailed examination. In particular, it assumes that (1) human error is entirely the result of misperception, misjudgment or inappropriate behavior on the part of the driver when some crashes that are coded as human error may in fact involve other factors that cannot be readily overcome by technology, and (2) that technology is error-free, or nearly so. Unfortunately, research on driver behavior and recent experiences with AD do not support either one of these assumptions.

2.2. Driver error does not mean driver culpability

Human error is a concept that is used in scientific crash analysis to identify and, where possible, categorize operator-induced causes, typically involving in-depth multi-disciplinary crash analysis. The decision to cite driver error as a crash cause by an officer or a crash investigator is often by default when there is no direct evidence of vehicle failure or inadequate infrastructure. A case in point is the determination of fatigue or falling asleep, where the estimated involvement rates vary from less than 5% to nearly 40% (Shinar, 2017). Unfortunately, in spite of new and widespread approaches to data collection, we have quite divergent estimates – often methodology-bound - of the extent to which many specific driver errors are actually involved in crashes (e.g., fatigue, speeding, and distraction).

It is important to emphasize that many crashes attributed to driver error may not be within the reasonable ability of the driver to prevent. Errors can arise from inadequate vehicle or roadway design (such poor design of controls and displays, visual obstruction from A-pillar, or inadequate sight distance or intersection configuration), and should more appropriately be labeled design-induced errors, but are attributed to the driver. In extreme cases, unintentional errors, such as misperception, can actually lead to subsequent events that make the crash unavoidable– a situation of an “accident waiting to happen” (see Hauer, 2016 for a fascinating discussion of such a situation involving bicycle lanes). It should also be noted that while automation can improve sensing, decision-making, and vehicle control it is not designed to eliminate deliberate violations that may be possible by drivers disengaging the automatic mode (that may result in reckless driving, speeding, not coming to complete stop at STOP signs, or passing in a no-passing zone).