Self-Driving Cars:
Risk Constellation and Acceptance Issues

Armin Grunwald*

Self-driving cars have become a challenging and discussed mobility option in Western societies in recent years. Technology is advancing quickly while simultaneously posing many ethical, legal, and social questions to the reflective scientific disciplines and to society as a whole. This paper focuses on the risk constellation of self-driving cars and draws some conclusions of their social acceptance. The final thesis is that an over-hasty introduction of self-driving cars motivated by economic competition might not only increase risk to road users but may also undermine the social acceptance of this technology. Hence, an ethical and legally responsible introduction should happen step by step in order to allow problems to be resolved as they emerge. Interdisciplinary cooperation between engineering, information technology, legal science, ethics, and the social sciences is needed to develop sound solutions to the many challenges of coping with risks and ethical issues of automated driving in a pro-active manner.

I. Self-Driving Cars – A Disruptive Innovation

The future of mobility is closely related to the ongoing digital revolution. In recent years, drivers have had access to more and more assistance based on advanced sensors, the real-time evaluation of the collected data, and actuators implementing conclusions made by algorithms. The latest processors and sensors are able to observe the traffic situation in the surroundings of a car in real-time and can determine the next steps to be taken in order to adapt the car to the respective traffic conditions. This development has already led to a partial automation of driving in new vehicles. Highly automated systems can autonomously change lanes and exert other functions without human intervention. In some countries, test fields have been established on which highly or fully automated vehicles can operate. In Karlsruhe (Germany), for example, a test field is currently being implemented by the Karlsruhe Institute of Technology (KIT) in an urban quarter. This test field will take into account the high complexity of urban mobility involving different participants in road traffic such as pedestrians, cars, bikers, and messenger and logistics services.

These developments put pressure on governments and society to reflect, to establish positions and to make decisions on whether and how autonomous mobility technologies could become part of the existing transportation system. Decisions have to be made whether and under what conditions automated driving systems can be approved, depending on ethical and legal issues. Questions of responsibility, accountability and liability have to be analysed and solved in order to create clear legal boundaries for self-driving cars, their developers, producers and owners as well as for mobility service providers such as car-sharing companies. Several ethical and even philosophical questions are involved, eg on the distribution of autonomy and responsibility between humans and technology by designing new interfaces between humans and technologies and possibly restricting human freedom to act as a driver in case of safer self-driving cars.

Autonomous driving is a fascinating field in a technological but also in ethical and legal respect.1 A huge wave of public interest has developed in recent years, with most of the attention being captured by the lat-

DOI: 10.21552/delphi/2018/1/7

* Prof. Dr. Armin Grunwald, Head of the Institute for Technology Assessment and Systems Analysis (ITAS) at Karlsruhe Institute of Technology (KIT) and Head of the Office of Technology Assessment at the German Bundestag. For correspondence: <arnim.grunwald@kit.edu>. The author builds on his own work (fn 4) and on the experiences in the ethics commission mentioned in fn 5

1 Markus Maurer; Jan Gerdes; Barbara Lenz; Hermann Winner (eds.), Autonomous driving. Technical, legal and social aspects (Springer Open 2016)
ter issues rather than by the technological developments. In these debates the frequently taken focus on simply replacing human drivers by algorithms and board computers is much too narrow. It is probably more important to recognise that autonomous driving opens up a wide space of new mobility options beyond the traditional individual auto-mobility with private cars; new mobility conceptions and patterns, new business models for mobility providers, and new combinations of private and public transport or even blurring the traditional borders between them, could become possible. This property only makes self-driving cars a possibly disruptive innovation.

Disruptive innovations, however, usually are disruptive in several respects. They might challenge the economy, existing business models, value-added chains, competition, consumer behavior, legal frameworks, incentive systems etc which can lead to major changes. A smooth and responsible transition to new transportation systems, including self-driving cars and other autonomous elements, requires proactive analysis as well as the exploration and development of feasible innovation strategies combining the various issues in an integrated manner. Interdisciplinary cooperation is needed between engineering, information technology, legal science, ethics, and the social sciences in order to develop sound solutions to the many challenges of coping with risks and ethical issues in a pro-active manner.

This paper focuses on the risk constellation of self-driving cars (Section II) and discusses some issues of acceptance (Section III). The final thesis is that a hasty introduction of self-driving cars motivated by economic competition might cause damage not only to human participants in road traffic and to people affected by accidents but also to the social acceptance of this technology. Hence, an ethically and legally responsible introduction should happen step by step in order to allow problems to be resolved as they emerge.

II. Risk Constellation of Self-Driving Cars

Technological advancement often changes societal risk constellations. Intended results frequently include, amongst other positive expectations, significantly improved safety standards, improvements in health, longer life expectancy, environmental efficiency and greater prosperity. However, technological innovations also often bring with them unintended and unforeseen consequences, including new risk types. In many respects, autonomous driving represents an attractive innovation for the future of mobility. Greater safety and convenience, use of the time required for driving for other purposes, better access to mobility for disabled and elderly persons, and efficiency gains on the system level are a few of the most commonly expected advantages. At the same time, the systems and technologies for autonomous driving are susceptible to risks. Comprehensive analysis and prospective evaluation of the possible risks of autonomous driving are an indispensable part of a responsible research and innovation process and thus equally important preconditions for acceptance both on the individual and societal levels.

Risk is possible harm that can occur as the result of human action and decisions. It contains three central semantic elements:

- **Epistemic moment of uncertainty:** What type and how severe is the harm that could occur? And how plausible and probable is it that the harm will actually occur? For both questions, the spectrum of possible answers ranges from scientifically attested and statistically evaluable to mere assumptions and speculation.
- **Ethical moment of the undesired:** Representing possible harm, risks are obviously undesirable. Nevertheless, the evaluation of the possible consequences of actions as a risk or opportunity may be disputed. One example is whether genetic modification of plants is perceived as a means of securing world nutrition or as a risk to humanity and the environment.
- **Social moment of risk distribution:** Often opportunities and risks are distributed differently among different groups of people. In extreme cases, the beneficiaries are not affected by possible harms at all, while those who bear the risks have no part in the expected benefits. It is crucial to consider who is affected by the opportunities and risks, in what ways and whether the distribution is fair.

A societal risk constellation consists of the relationship between groups of people such as decision-makers, regulators, stakeholders, affected parties, advi-

---

2 Armin Grunwald, Technology Assessment in Practice and Theory (Routledge 2019) 16 ff
sors, politicians and beneficiaries in view of the frequently controversial diagnoses of expected benefits and feared risks. It is among the tasks of technology assessment\(^3\) to make the respective risk constellation of new technologies as transparent as possible. Metaphorically speaking, a map of possible risks and groups affected has to be developed in order to inform public debate and political judgment, thereby taking into account all the three risks mentioned above. Autonomous driving shows several possible types of risk, of which some are well-known from traditional driving while others are new, often related with the digitisation of driving. The following risk types have been identified in the framework of a project of the Daimler Benz Foundation.\(^4\)

(1) Individual accidents: One of the expected advantages of autonomous driving is a major reduction in the number of traffic accidents and their consequences in terms of harm to life, health and valuables because human misbehavior is the reason for 90 to 95% of road traffic accidents. However, due to technological problems or in situations for which the technology is not prepared, accidents specific to autonomous driving can occur. Two accidents with fatalities have already occurred in the United States. Accidents are omnipresent in the practice of everyday traffic. The fact that these accidents happen is widely accepted in society. This is exemplified by the fact that the currently more than 3,000 annual traffic deaths in Germany, for example, do not lead to protests or public rejection of car transportation. The gradual introduction of self-driving cars into the transportation system will provide the chance to learn from incidents and make step-by-step improvements. Continuous monitoring and careful cause analysis will be of critical importance here. In particular, accountability and liability issues must be investigated and clearly determined, based on a clear distribution of responsibilities among car developers and producers, mobility service providers, the owners of self-driving cars, and possibly further groups involved.

(2) Systemic risks: Autonomous driving adds new types of systemic effects to the existing ones. While in the automotive world to date vehicles are operated more or less independently of one another and mass phenomena only occur through the unplanned interactions of the individually guided vehicles, autonomously guided traffic will to some extent be connected through control centers and networking. Through the control software and the reliance on the internet, new effects could emerge. The control of a large number of vehicles will in all likelihood be conducted through software that is identical in its fundamental structure, as the complexity and concentration of companies will presumably strongly limit the number and diversity of providers. This situation could lead to the simultaneous breakdown or malfunctioning of a large number of vehicles based on the same software. Naturally this leaves the system vulnerable and a number of security issues must be taken into account. Software hackers might influence and damage the mobility system which is not present to date. The system may also be vulnerable to attacks by terrorists, mentally ill individuals or even militarily motivated cyber warfare. Control centers could be hacked, malware installed, or even a system collapse triggered through malicious action. These problems would then no longer remain on a micro-scale but could take on economically significant proportions and could hit a large number of people simultaneously. This is primarily an issue of software dependability and IT security.

(3) Investment risks: Research and development of autonomous driving is extremely cost-intensive. Considerable additional investment is required before any introduction of autonomous driving to the marketplace is possible. As with other investments, the business risk exists that the return on investment may not be on the expected scale or in the expected timeframe due to autonomous driving failing to catch on on a large scale, e.g. because of low acceptance by users. Even after a successful market launch, mishaps or technology-related system effects can occur posing a major risk for the affected brands. Special attention in this context needs to be paid to the risk factor posed by the complexity of the software required for autonomous driving. Complex software is impossible to test in its entirety. Unexpected problems can occur in actual use as we know from any other software. However, self-driving cars behaving

---

3 Armin Grunwald, Technology Assessment in Practice and Theory (Routledge 2019)
4 Armin Grunwald, ‘Social risk constellations for autonomous driving, Analysis, historical context and assessment’ in Markus Maurer, Jan Gerdès, Barbara Lenz and Hermann Winner (eds) Autonomous driving. Technical, legal and social aspects (Springer Open 2016) 641-662. This part follows closely the line of argumentation developed there
unexpectedly due to software problems may create risks to other traffic participants or bystanders. When an autonomous vehicle causes an accident due to a software error, it is unacceptable for ethical and legal reasons. In such cases the massive media attention which would follow could undermine trust in the products of a specific car company and create massive economic problems.

(4) **Labor market risks**: Automation is associated with concerns about the loss of jobs in many fields of work. Comprehensive introduction of autonomous driving would undoubtedly affect the labor market. The primary losers would be drivers of vehicles which are currently manually operated: truck drivers, taxi drivers, employees of logistic and delivery companies. A mobility system completely converted to autonomous driving could in fact largely do without these jobs altogether. On the other side, highly qualified personnel will be needed in the development, testing and manufacture of the systems, particularly in the supply industry. New jobs to operate new business models and new mobility services could also emerge. But even if there was a balance regarding the amount of work to be done by humans, the issue of distribution will remain; the elimination of unskilled jobs and the creation of new, highly skilled positions will lead to a field with winners and losers as well. This potential problem needs early political consideration.

(5) **Privacy risks**: Even today modern automobiles leave electronic traces, eg by using navigation aid or through data transmitted to the manufacturer. If autonomous vehicles are connected to the internet at all times, the electronic trail would amount to a complete movement profile. Movement profiles provide valuable information for intelligence services, which could for instance track the movements of regime opponents, but also for companies, who could use such information to create profiles for targeted advertising. In view of the increasing digitisation and connectedness of more and more areas of our social and personal lives, the specific additional digitisation in the field of autonomous driving would presumably represent just one element among many others. The problem is immensely larger, as current debates about Big Data and many scandals such as the Facebook Cambridge Analytica case show. While autonomous driving is presumably of only minor specific significance in this debate it is of greatest significance to democracy to solve the problem at large.

(6) **Dependency risks**: In the case of a large-scale shift of mobility capacity to autonomous driving, a high proportion of society’s mobility needs would naturally depend on the functioning of this system. A breakdown, due to software problems or cyberattacks (see above), would be manageable only if there were enough people who could still operate the vehicles manually. If a large share of the logistics and freight traffic were switched over to autonomous systems, it would be impossible, in the case of a longer-term total breakdown of the system, to maintain a sufficient pool of drivers, not to mention the fact that the vehicles would have to be equipped to enable manual operation in the first place. Even if significant logistics chains were to be interrupted for a lengthy period due to a system failure, bottlenecks could still quickly form, both in terms of supplying the population and maintaining production in the manufacturing industries. The ongoing digitisation of mobility will further increase society’s vulnerability to intentional disruptions and external attacks. Standby or backup measures to address these issues are necessary and technically feasible.

The risk constellations described here are based on qualitative and exploratory considerations from the current world of mobility. They therefore have a certain plausibility, but also involve speculative aspects. They should not be understood as predictions, but as guideposts that should be observed along the way to the research, development and introduction of autonomous driving. The evidence suggests that a responsible introduction of this technology is possible. However – would people accept or even use and love self-driving cars?

### III. Introduction Scenarios and Risk Perception

Many concerns arise around the issue of public acceptance of autonomous driving. The so-called ethics dilemma (the trolley problem), the unclear distribution of liability among several actors in the field, challenges to the distribution of responsibility be—
between autonomous technology and humans, uneasiness with accepting an on-board computer making decisions on life and death, and psychological issues of boarding a car without a human driver and trusting the machine to safely navigate through today’s complex traffic are among the issues that could prevent the willingness of possible users to change their mobility behaviour.

Acceptance cannot be manufactured but can only develop (or not). Its development depends on many factors, some of which can certainly be influenced. First of all, public as well as individual acceptance depends largely on perceptions of the benefits and risks of self-driving cars. It is crucial that the expected benefits manifest themselves at the individual level of users. Furthermore, it is essential that the involved institutions (manufacturers, operators, regulators, monitor and control authorities) enjoy public trust. Therefore, communication about possible risks must be conducted in an environment of openness – nothing is more suspect from a public debate standpoint than to assert that there are no risks and that everything is under control. The history of nuclear power in many European countries is an excellent example of this. Instead, concerns and questions of users and citizens must be taken seriously. All of this requires early and open communication with relevant civil society groups and citizens as well as in the mass media sphere.

The public perception of the risk will, in addition, depend largely on how autonomous driving is introduced. If it happens as part of a gradual automation of driving, the potential to learn gradually from the experiences gained along the way will greatly lessen possibilities of diagnosing autonomous driving as a high-risk technology for passengers and bystanders. Unlike switching on a nuclear reactor, for example, the process of increasing driver assistance towards greater automation has so far progressed gradually. We are comfortable with ABS, ESP and parking assistants. Incremental introduction allows for a maximum degree of learning and would also enable gradual adaptation of the labor market, for example, or alleviation of privacy concerns (see above).

In more revolutionary introduction scenarios, risk perception could be faced with more radical and fundamental issues. The public perception then could react more sensitively to accidents or critical situations. Consequently, the risk of ‘scandalisation’ would be greater and investment risks (see above) could develop into a real problem for individual suppliers or brands in case of severe accidents in line with a hasty introduction. Hence, an ethical and legally responsible introduction should happen step by step in order to allow learning from problems as they emerge. However, this approach also includes risk. A gradual shift from conventional to automated driving would imply a long time of co-existence of both types with increased complexity. Also the well-known dilemma of automation has to be taken into account.6

There is some reason to believe that for the acceptance of autonomous driving, expected benefits will outweigh concerns regarding risks.7 A focus on risks would therefore presumably miss the core of the challenge: the decisive factor seems to be the expected benefits not at the macro-economic but at the individual level of end-users. Rather than focusing on risk, it seems appropriate to regard the elements and options of autonomous driving as parts of an attractive mobility future with greater safety and efficiency, more social justice and more convenience/flexibility. Of course, there is no zero-risk scenario – but that has not been the case with conventional driving either.

A big unknown is human psychology. Whether, to what degree and under what conditions people will entrust their lives and health to self-driving cars is an open question. Other autonomous transportation systems such as subways or shuttle services did not face acceptance problems. But railway vehicles are perceived different from cars, in particular because of much lower complexity of the system. Also the controllability of risks is different. Therefore experiences from autonomous railway services cannot be simply transferred to road traffic.

Another aspect with regard to acceptance is that in conventional automobile transportation a well-developed culture of damage adjustment is in place through the traffic courts, appraisers and insurance companies which has reached a high degree of precision and reliability. Autonomous driving, by con-

---


7 Armin Grunwald, ‘Social risk constellations for autonomous driving. Analysis, historical context and assessment’ in Markus Maurer, Jan Gerdes, Barbara Lenz and Hermann Winner (eds), Autonomous driving. Technical, legal and social aspects (Springer Open 2016) 656 ff
trast, would pose new challenges to the damage adjustment system. Questions such as ‘Who caused the damage, man or machine?’ or ‘Who is liable, the mobility service provider or the manufacturer?’ have to be answered in an unambiguous and legally unassailable manner. Acceptance of autonomous driving will also depend on the development of adequate and clear answers to these questions.