Thinking outside the box: Inventivio Innovation Briefs
Issue 2016-09

Top misconceptions of autonomous cars and self-driving vehicles

Alexander Hars, Inventivio GmbH

ahars@inventivio.com

Self-driving cars are a rapidly evolving technology which only a few years ago was still considered science fiction. In such a dynamic context, quick intuitions can be very misleading and misconceptions about the technology, its impact, and the nature of the innovation process abound. In the following we address some of the most widely held misconceptions about autonomous vehicles.

This list of misconceptions can be used to :

- enlarge your knowledge about the underlying forces and paths that will shape the future of autonomous vehicles, and to
- assess the expertise of authors and experts who are making statements about self-driving cars.

You will find that many widely repeated statements about autonomous vehicles can be attributed to very narrow perspectives on self-driving cars and a lack of understanding for the nature of the global, distributed innovation process which drives this technology forward.

Note that this article focuses on fully autonomous vehicle technology, i.e. vehicles that can drive themselves without human intervention, even empty, without any human in the car.

1. Driver assistance systems will evolve gradually into fully autonomous cars

This is an extremely attractive misconception that you will find repeated over and over. At first glance, it seems to be very logical and rooted in history: If we look at the past 10 years as well as the present we find that each new car model comes equipped with more computational power, more electronic safety and assistance features – from auto-parking, lane warning, intelligent cruise control, to emergency braking etc. Isn't it natural to extrapolate this trend into the future?

But this evolution contains one obvious discontinuity: All of the driver assistance systems which are in use today operate only for short times and in extremely limited settings. Auto-parking operates for a few seconds with the driver watching. Emergency braking kicks in at the last moment before an inevitable crash. Lane warning comes on briefly when a car veers out of its lane.

This changes drastically once the car drives itself **continuously** for minutes or hours. Here, gradual evolution is impossible: from the moment that a car drives continuously, there is no margin for error; no room for gradual improvement, learning by doing or evolution. It needs to be able to cope with all short-term eventualities and crisis situations that may arise on the spot.

People often argue that such assistance systems need to be supervised by the driver. This makes sense for assistance systems that operate for a few seconds or minutes (such as a parking assistant) but it can not work for systems that drive continuously. Humans are not capable to maintain the state of alert for hours and hours which would be required to immediately counteract possible deficiencies of a driver-assistance system or to take over from it in a split-second.

We can only entrust the driving task to a driver assistance system when we are sure that this system can handle all situations which arise suddenly and require immediate reaction. This means that driver assistance systems operating continuously on a highway need to be able to cope with rare situations including pedestrians and bicyclists on the highway

(they do appear sometimes on highways), accidents unfolding, animals, sudden rainfall etc. Gradual evolution of such systems is impossible; they need to be extremely capable from the first day on which they are put into operation.

If we systematically enumerate the risk scenarios which a continuously operating driver assistance system needs to handle, we find that it must pass almost every risk scenario that a fully autonomous vehicle must be capable of handling. Only those few scenarios that do not arise suddenly but that can be anticipated minutes before (e.g. that the exit is coming up) need not be handled by a continuously operating driver assistance system because it can alert the driver in time and request to return control to the driver.

To summarize: Driver assistance systems can not evolve continuous driving capability gradually! At the moment we entrust them to drive continuously they require a huge, discontinuous jump in capability which will place their capabilities very close to the capabilities of fully autonomous vehicles.

2. The first models of fully autonomous cars will be targeted to the consumer and will be available for purchase

When will I be able to buy an autonomous car? When will autonomous vehicles appear on the market? These key questions already contain an innocent assumption about the market for self-driving vehicles: that these cars will be primarily targeted to the consumer.

Unfortunately this assumption disregards both the difficulties and opportunities associated with fully autonomous cars. A key problem is the region where cars are capable of autonomous operation. Consumers who want to purchase a car expect it to operate in most parts (at least all highways) of the country or ideally the whole continent, and preferably in all non-extreme weather situations. This is a tall order! Detailed maps need to be created and maintained; algorithms need to support more than dry weather and light rain but also snow and heavy rain.

For auto manufacturers this means that solutions need to be found that essentially work on the entire planet. The structure of the maps needs to be defined and then the maps themselves need to be collected. This will be a major task because the maps need to be much more detailed than the conventional street maps currently being maintained by Google, Apple, TomTom, Nokia Here and others. It is not yet clear what the best structure for such a map is (and nobody has even started to address the problem of how to navigate in snow-covered areas which may in turn have implications on the mapping approach). Therefore a significant lead time will be necessary before an auto manufacturer can release models to the market that are capable of driving autonomously in most parts of a country or even continent.

This is not a problem for the other use case of autonomous vehicles. Fully autonomous cars can operate much earlier on a limited set of predefined routes as taxis or buses which provide mobility as a service. Consumers will not be very interested in such autonomous vehicles but taxi companies, Uber, Car2Go, car-sharing and rental car companies as well as transit corporations clearly see the potential of autonomous vehicles: These autonomous cars can provide local mobility as a service at low cost. In urban areas fleets of autonomous cars can be called by anyone via mobile app. A self-driving taxi will arrive a few minutes later and drive the passenger to their destination with maximum convenience and without the need to look for parking. The autonomous taxi will drop the passenger of at the target location and continue on to the next customer. The business potential of such fleets is enormous as we have shown in other papers because such fully autonomous mobility services will be able to provide local mobility at much lower cost per vehicle-kilometer than today's individually owned vehicles and today's taxis.

Mapping just an urban area for autonomous driving (and keeping the maps up to date) is a much smaller problem than mapping the whole country. Additional problems of autonomous driving which delay the introduction of fully autonomous cars on a nationwide scale can be circumvented: Current autonomous cars can operate only in sunny areas with little rain and without snow. In the US alone there are hundreds of cities which fit this profile and where fleets of autonomous cars can operate safely long before the harder problem of autonomous driving in very adverse weather is fully solved.

The business case for fleets of autonomous cars is strong. Because such fleets are most attractive if they have many customers (and cars) in a region, such fleets exhibit network effects: First movers can achieve a much better market position than followers. In this market, the market leaders will be much more profitable than their competitors. Therefore entrepreneurs and investors have an incentive to grab market share early. A strategy where fleet operators deploy autonomous taxis in regions without adverse weather first with the goal to expand a few years later into the rest of the country – when the problem of operating autonomous cars in snow is solved – is therefore much more attractive than waiting until autonomous cars are capable of operating everywhere.

Fleets have another key advantage for deploying the first fully autonomous cars: Fleet owners maintain full control over their cars. Whereas cars sold to a consumer end up dispersed across the country, fleet owners continue to have access to the physical car at all times. When problems arise, accidents happen, updates or maintenance are required, fleet operators can easily access all of their autonomous cars. This is very important in the early phases of autonomous vehicles because operators of autonomous vehicles that operate in a limited local region can then remotely monitor all their cars and dispatch service teams where needed.

This level of control reduces risk: Fleet operators can ensure standard levels of service and maintenance whereas this is much more difficult for vehicles delivered to customers across the country. The risks of hacking and tampering with the car can also be also reduced when the cars are under control of a fleet where physical access to the car's innards is

limited to authorized personnel. It will be much easier for hackers to gain access to an autonomous vehicle which they can purchase than to a fleet vehicle.

Finally, vehicles that operate in urban areas can productively operate at lower speeds than cars sold to the consumer. Even at a maximum speed lower than or up to 50km/h (~35mph) local autonomous taxis can be very productive. Although urban autonomous driving is often considered the most difficult context, the liability risks of highway driving can be much higher because of the higher speed. In urban settings, cars can come to a stop when a problem arises in front of them. On highways inadvertent braking just to be on the safe side can cause major accidents and damage. The ability to operate autonomous fleets in select areas initially with low speeds at low risk profiles is another reason why autonomous vehicles are more likely to appear in fleets first and will only later be available for purchase by consumers.

In summary, these issues clearly show that we should not expect the first fully autonomous cars to be available for sale. The first fully autonomous vehicles are much more likely to appear within fleets of autonomous taxis or buses that operate in select urban regions. It may then take several more years until the first autonomous vehicles become available for purchase.

3. It will take decades until most of the vehicles on the road are capable of autonomous driving

It took automotive innovations such as anti-lock braking, airbags, seatbelts several decades from being introduced in cars of the premium segment until they trickled down to all models and until most of the cars on the roads were equipped with these innovations. Many people assume that autonomous driving will also take this long. But there are good reasons why the diffusion process for self-driving cars (though not for driver assistance systems) will adopt a different pattern:

The slow diffusion of classical automotive innovations results from the limited additional benefit of these technologies. It is difficult for a buyer to quantify the benefits of anti-lock braking or airbags in dollar terms and to determine whether they exceed their often initially hefty price. Severe accidents – in which these technologies would make a difference – are rare and therefore it may be a rational decision for many, not to purchase the new technology. The higher the buyer's wealth and the higher the buyer values his life (both of which are correlated), the higher the likelihood that someone will adopt such an innovation. This is why these safety-related innovations are so strongly tied to the premium segment in the beginning. As the technology is adopted more and more, their costs gradually sink which means that the cost-benefit calculation turns positive for more and more prospective buyers.

Deploying early self-driving cars in locally restricted fleets can reduce operational, weather and hacking risks

But increased safety is not the only key benefit of autonomous vehicles: Self-driving cars unleash the driver from the steering wheel and thereby increase available time – a precious and scarce resource. The benefits are most obvious in logistics, where the costs for a truck driver represent about a third of the total transport costs (Bayliss, B.T, 2012, p. 14). Fully autonomous technology dramatically increases the return on investment and will therefore lead to rapid adoption. Consumers also value their discretionary time. If an average driver spends about 1 hour per day behind the wheel this translates to a 15 full days of additional time gained each year! Someone who values their own time at only $5 ($10) per hour, values the time-gaining benefit of this technology at $1,825 ($3,650) per year. If the buyer intends to keep the vehicle for 5 years, then he would be willing to pay $9,125 ($18,250) for this feature. Given these benefits, the group of early adopters would be large and not at all limited to the buyers of premium cars. The more time people spend in their cars, the higher the incentive to purchase a fully autonomous vehicle. In contrast to the increase in safety associated with many classical automobile innovations, which can only be expressed as a reduction in the probability of an accident and associated damages, the benefit of additional discretionary time is a solid fact which the buyer can be certain to experience every time he uses the car.

Another solid benefit of self-driving vehicles are lower insurance premiums. Because self-driving cars will be much safer, the premiums will shrink. However it may take a few years for this effect to kick in. Insurance companies first need to be able to quantify the risk associated with fully autonomous vehicles. Young drivers – who pay the highest premiums today – will benefit most from the lower premiums. Understanding the direction that the technology is taking and being concerned about the risks of manual driving, many parents will decide not to spend the money for driving lessons and guide their children toward fully autonomous driving instead.

Fully autonomous vehicles can be used as self-driving taxis. In sufficiently populous areas, fleets of autonomous cars will emerge that provide mobility as a service. These cars can provide mobility at much lower cost than privately owned vehicles because their utilization rate will be much higher than the dismal utilization rate of approximately 5 percent for privately owned vehicles. Because of higher utilization rates, these fleets are much less sensitive towards the cost of autonomous technology and can adopt the technology much earlier than privately owned vehicles. In cities many privately owned cars will be displaced by such fleets. Several simulation studies have shown that each self-driving taxi may replace 6 to 10 privately owned vehicles. Because such fleets exhibit network effects and tend to become monopolies, there will be an intense fight for market leadership. This will accelerate the diffusion of self-driving taxis and will rapidly increase the

number of miles traveled in autonomous cars.

Besides individuals and fleet operators, governments play a key role to ensure that this technology will be adopted with unusual speed: Self-driving cars are a critical, enabling technology which will impact almost every sector of industry. New business models become possible; in addition, transportation costs for people and goods will shrink considerably. Countries which are slow to adopt this technology may loose their competitive positions; early adopters may significantly improve their competitive advantage. Therefore many governments will work hard to ensure that they are well positioned with respect to this technology. Such patterns are already visible in the United States, where the competition between states for leadership in autonomous car technology has been in full swing since 2012 when Nevada became the first state to pass an autonomous vehicle law. After intense lobbying California followed suit; many other states have sinced worked on autonomous vehicle legislation – not all successfully. Michigan is an example for a state that is very concerned about loosing its position as heart of the US auto industry to California and therefore has also passed a law for autonomous vehicles alongside with additional measures aimed to increase Michigan's competitive position in this emerging technology. In Europe, the United Kingdom has recognized the technology's potential and is investing hundreds of million of pounds to grow an autonomous mobility industry. The UK has not signed the Vienna Convention on Road Vehicles and therefore can more easily introduce autonomous cars on their roads. The first such projects – most notably in Milton Keynes were 40 autonomous pods will ferry passengers from the train station to the city center, are being implemented. In Asia, Singapore looks very actively to improve local mobility through driverless car technology. It is only a question of time until China recognizes the potential of this technology for reducing congestion and pollution in their cities via autonomous vehicle fleets as well as reducing the size of required investments in their road infrastructure – much of which is still being built – and makes this technology a top national priority.

4. Self-driving cars are controlled by classical computer algorithms (if-then rules)

When thinking about self-driving cars, we easily fall into the trap of envisioning them as controlled by huge programs consisting of millions of elaborate rules which specify how the car must act for every possible situation. But a self-driving car does not have any rule such as:"IF a ball rolls across the road THEN reduce the speed and watch for children that might come running after the ball." Programmers who develop a self-driving car do not work with some exceptionally detailed "theory of driving" which they then translate into computer algorithms.

Such cars heavily rely on machine learning and pattern recognition – approaches from the field of artificial intelligence. A core problem, for example, is interpreting sensor- and image data: What kind of an object is in front of the vehicle? Is it another vehicle, a pedestrian, a bicycle, an animal? Instead of developing many rules for recognizing these objects, a sophisticated learning algorithm is fed with many images containing objects. Each image is annotated with the kind of object that it contains. Now the learning algorithm begins its work. It examines the images and guesses the kind of object in each image. Initially most of its guesses will be wrong. Therefore the algorithm modifies internal parameters or parts of its structure somewhat and tries again. This process continues, discarding changes that reduce the algorithm's accuracy, keeping changes that increase the accuracy, until it correctly classifies all images. Afterward, when entirely new images are presented to the algorithm it will classify them with high accuracy. The algorithm has learned! Thus a driverless car may not explicitly recognize the ball but rather consider it as an unidentified object which is unexpectedly moving across the road. It does not need to increase its vigilance (which is already at 100%) to search for children which might follow the ball but it will treat this as a disturbance of the road situation by an unidentified moving object which implies an additional risk and therefore – depending on the complexity and clarity of the road situation – reduce its velocity or stop altogether. The precise form of the object is not important – if a lone skateboard were to suddenly slide across the street the car would react in the same manner.

The learning approach can also be used for actions and evaluations. Instead of supplying the vehicle with a fixed evaluation scheme from which the right action for each situation can be deduced, the programmers feed the software with many traffic situations and specify the correct action for each situation. The program then searches by itself for the best configuration of internal parameters and internal decision logic which allow it to act correctly in all of these situations. Like with us humans, it then becomes difficult to answer the question why the car exhibits a specific behavior in a new situation: no "explicit rules" have been specified; the decision results from the many traffic situations to which the algorithm had been exposed beforehand.

Therefore self-driving vehicles are not programmed in the classical sense; they need to learn. It is not possible to reduce human driving decisions to a few (not even very many) IF-THEN rules. The development of autonomous vehicles is not only a challenging software development problem. It requires an extensive learning strategy where vehicles are exposed to a huge number of traffic situations. Google, as a consequence, has driven almost two million kilometers on public roads with test drivers and has assembled an enormous fund of traffic situations from which its vehicles can learn.

Another characteristic of self-driving cars is their use of probabilistic reasoning. For example, a car does not assume that it knows its exact position. Instead it maintains a distribution of positions at which it currently might be with certain probabilities and – somewhat simplifying – takes the position with the highest probability as its current position. Another unusual aspect of some algorithms used in self-driving cars is the use of randomly generated numbers in the decision-making process. As with hu-

man decision making, this means that a self-driving vehicle may behave differently when faced with exactly the same situation multiple times!

To summarize, we should avoid conceptualizing self-driving vehicles as machines which are controlled by a detailed, exactly specified and in principle comprehensible software program. Instead we should conceptualize their behavior as being the result of a long and varied program of learning. The capability of such cars can be analyzed through simulation and testing but not just by examining its source code.

P.S.: Please avoid referring to self-driving vehicles as 'automated' vehicles. This term multiplies the misconception that these cars – like all the other automated machines in our factories – operate by following very clearly described step-by step rules which were programmed into them one by one. Our current notion of 'automation' brings up the wrong intuitions about how these cars operate.

5. Public demonstrations of self-driving cars provide an indication of their capabilities

Rarely does a month go by without a high-profile public demonstration of a self-driving car: In Spring 2015 a self-driving Delphi car completed a 3000 mile journey from the West Coast to the East Coast of the United States. Earlier an autonomous Audi had raced at maximum speed around a German test track. A year before a self-driving Mercedes had driven a 100km route on public roads in Southern Germany. Such events provide great publicity. They capture the public's imagination but at the same time they easily mask fundamental shortcomings from which these prototypes still suffer.

Autonomous driving technology has advanced greatly over the last 10 years since the DARPA's Grand Challenges in 2004, 2005 and 2007 kick-started the development of self-driving cars. Today, small teams from a university, auto manufacturer or auto suppliers can build a basic autonomous car in just a few months which is capable of driving in sufficiently mapped areas, detecting traffic signs and stop lights, detecting other vehicles, staying in lanes, accelerating and braking smoothly etc. When such a prototype drives a route on public streets, it is hard to avoid the impression that the future is already here!

But enormous challenges remain until self-driving cars are ready to drive on public streets without test driver. Such cars need to be able to cope with almost every situation – however remote – which they may encounter. They not only need to correctly detect and categorize objects in many situations, they also need to anticipate their behavior. The problem space is large and very complex. Although it is possible to identify a large number of possibly problematic situations using risk analysis it is impossible to enumerate all potentially problematic situations.

Therefore much of the work for developing a mature self-driving car consists of covering a large number of rare road scenarios and making sure that such scenarios can be handled. Because these scenarios occur rarely, public demonstration do not allow the observers to distinguish between a quickly developed prototype and a mature self-driving car.

Besides working diligently through thousands of risk scenarios, self-driving cars need to be exposed to actual driving situations. Testing them on millions of kilometers is vital to ensure that risks which are hard to anticipate can be detected. Google makes an enormous effort to validate their autonomous driving technology and to gain experience via extensive testing on public roads. Their test cars are usually staffed by two persons while testing: A test driver ensures that the car operates safely and is ready to counteract any mistakes that the car might make. The second person constantly monitors the car's algorithms and records any issues or interesting problem situations which might warrant further analysis and perfection.

The early phase of Google's self-driving car project can be used as an example for the kind of problems that can only be detected through intensive testing on public roads: The Google prototype car was driving autonomously on the highway. It had rained shortly before and truck tires occasionally sprayed water high into the air. This led to problems because the car interpreted the spray as a solid object which unexpectedly got into the way. It was not hard to solve this problem but it is an example for issues that are difficult to anticipate and that must be found out by live testing.

To summarize: It is impossible to judge the maturity of a self-driving car by observing public demonstrations. Difficult situations don't occur that frequently and therefore these demonstrations can only confirm that a prototype has reached quite a basic level of capability. The enormous difference in maturity between, for example, Google's prototypes – the current leader in this technology with nearly two million kilometers of testing in autonomous mode and more than 10.000km of testing being added every week! – and the prototypes of all other developers of autonomous car technology can not be appreciated by observing public demonstrations.

6. Self-driving cars need to make the right ethical judgements

As an autonomous car drives down a street, a frail old man suddenly steps into its path from the right. Simultaneously, a child steps into its path from the left. It is too late to brake. If the car swerves to the right, the old man dies, the child lives. If it swerves to the left, the old man lives, the child dies. If it continues straight ahead both will die. What is the ethically correct decision for the car?

Variations of this kind of ethical dilemmas – often referred to as the 'trolley problem' currently receive much attention. At first glance this seems to be a really difficult question. In the following we show that the problem is largely irrelevant for self-driving cars. We progress from weakest to strongest arguments:

a) No good solutions to these dilemma exist or can exist. Humans are not able to make a 'right' choice when faced with such situations either.

This dilemma is a good starter for night-long discussions. None of the alternatives one comes up with is 'ethically right'. If a human driver is in the same situation, he will necessarily make a choice but any action he chooses, is a bad action. How can we require a machine to make an ethical choice that no human is capable of making?

b) These dilemmas assume certainty and knowledge that does not exist in such situations.

For these dilemmas to work, the harmful outcomes for all of the actions must be known and certain. But in practice, nothing is certain. There is no certainty about the extent of damage for each of the actions. There is no certainty about the behavior of either of victims as the car approaches them. These cars can not have exact knowledge about the age, gender, health etc. of the persons in front of them, and can not correctly predict the resulting harm.

c) These dilemmas are always incredibly contrived. The probability that a car faces such a situation is extremely low.

Why don't we discuss such dilemmas today where billions of trips are being taken daily in cars and several thousand people die each day in traffic accidents? Cases where drivers face such situations are extremely rare today and may be even less probable for self-driving cars. From a practical perspective, therefore, these dilemmas may be completely irrelevant.

d) The question is wrong.

When looking at ethical questions there can be a huge difference between considering what is right and considering what is wrong. The ethical dilemma is usually presented in such a way that the self-driving car needs to take the ethically 'right' decision. As we know, the trolley problem has no ethically right solution- because in principle we can not weigh one life against another – , which makes it practically impossible for self-driving cars to solve the dilemma.

But – like humans who face this problem – self-driving cars do NOT need to adopt ethically right decisions. Our legal system and our ethics have evolved sufficiently to realize that many problems exist where it is hard to decide whether an action is legally or ethically correct. The standard by which we measure actual behavior against the law and against our moral compass therefore is not so much whether an action is ethically right but rather whether an action is ethically wrong: Actions must not violate laws or ethical standards! This difference in the problem statement matters! Instead of requiring self-driving cars to positively take ethically correct decisions, what our society really requires of them is that they avoid making ethically wrong decisions!

If we reformulate the dilemma in this way, the fundamental problems vanish. Neither is it right to kill the child nor is it right to kill the old man. But as it is impossible to avoid one of these outcomes, neither action can be characterized as being legally or ethically wrong. While both outcomes are bad and deplorable no court would find the algorithms at fault because they led to one or the other harmful actions. Exactly because there is no 'right' decision that either the victim must be the child or the victim must be the old man, no court will argue that the actual action taken by the self-driving car in this specific scenario was wrong.

In summary, much of the current discussion about the ethical dilemmas of life and death decisions related to self-driving cars is misplaced because it is concerned with finding right decisions where no right decisions are possible instead of realizing that self-driving cars can get by as long as they are able to avoid making decisions that are wrong.

7. To convince us that they are safe, self-driving cars must drive hundreds of millions of miles

One of the most difficult questions for self-driving cars concerns their safety: How can we determine whether a particular self-driving car model is safe? The most popular answer to this question is based on a straightforward application of statistics and leads to conclusions such as that "...fully autonomous vehicles would have to be driven hundreds of millions of miles and sometimes hundreds of billions of miles to demonstrate their reliability...". This statement comes from a recent RAND report by Nidri Kalra and Susan Paddock on the topic. Unfortunately, these statements are untenable in this form because the statistical argument contains major oversights and mistakes, which we will point out in the following.

7.1 Failure rate estimation

The argument is usually presented as a problem of failure rate estimation where observed failures (accidents involving self-driving cars) are compared against a known failure rate (accident rates of human drivers). Accidents are modeled as discrete, independent and random events that are determined by a (statistically constant) failure rate. The failure rate for fatal accidents can be calculated by dividing the number of accidents with fatalities by the number of vehicle miles traveled. If we consider the 32,166 crashes with fatalities in traffic in the US in 2015 and relate them to the 3.113 billion miles which motor vehicles traveled, then the failure rate is 32,166 / 3.113 billion = 1.03 fatalities per 100 million miles. The probability that a crash with fatality occurs on a stretch of 1 mile is extremely low (0,0000010273%) and the opposite, the success rate, the probability that no accident with fatality occurs on a stretch of 1 vehicle-mile-traveled (VMT) is very high (99,999998972%). By observing cars driving themselves, we can obtain estimates of their failure rate. The confidence that such estimates reflect the true failure rate increases with the number of vehicle miles traveled. Simple formulas for binomial proba-

bility distributions can be used to calculate the number of miles which need to be driven without failure to reach a certain confidence level: 291 million miles need to be driven by a self-driving car without fatality to be able to claim with a 95% confidence level that self-driving cars are as reliable as human drivers. This is nearly three times the distance between fatalities that occur during human driving. If we relax the required confidence level to 50%, then at least 67 million miles need to be driven without fatality before we can be confident that self-driving cars are safe. Although this calculation is simple most authors - including the authors of the RAND report - use the wrong measures. Instead of dividing the number of crashes involving fatalities (32,166) by VMT, they divide the number of fatalities (35,091) by VMT. This overstates the failure rate of human drivers because a single accident may lead to multiple fatalities and the number of fatalities per fatal accident may depend on many factors other than the reliability of the driver.

Accident type	Miles needed at confidence level	
	95%	50%
With fatality	291 million	67 million
With personal injury	5.4 million	1.2 million
Crash (=police-reported accident)	1.5 million	344.000
Any accident (estimated)	745.000	172.000
Safety critical incident (see 7.9)	(Human reliability info n.a.)	

7.2 Correlations matter!

Most authors duly proceed to calculate the failure rate for other types of incidents: The 1.715 million crashes with personal injury lead to a failure rate (including fatalities) of 56 failures per 100 million VMT. At a confidence level of 95% (50%) fully autonomous vehicles must travel 5.4 (1.2) million miles without failure. If we add the 4.387 million police-reported crashes without injury this leads to a failure rate of 201 failures per 100 million VMT. To reach a confidence level of 95% (50%) fully autonomous vehicles must travel 1.5 million (344,000) miles without failure. The RAND report points out that many accidents are not reported. The relationship of actual to reported accidents is not clear; some scientific studies have shown that it may be in the range between 2 and 4. The US Census uses the factor of 2 to estimate the number of accidents from the number of crashes. If we take the census estimate, then self-driving vehicles would have to travel only 745,000 (172,000) miles without accident at a confidence level of 95% (50%) to conclude that self-driving cars are as reliable as cars driven by humans.

If we consider different types of failures, not just accidents with fatalities, then the number of miles which need to be driven drops greatly. This brings us to the largest flaw in the statistical argument: the probability distributions of accidents with fatalities, accidents with personal injuries, and other accidents are correlated, they are **not independent**! A car that is much better at avoiding non-fatal accidents than human drivers is also likely to be better than human drivers at avoiding fatal accidents! A self-driving car that has a high rate of fatal crashes will also have a high rate of non-fatal crashes and vice versa. Because we also have other measurable failure rates besides the fatality rate such as the crash rate, the personal injury rate and the overall accident rate, we can gain insights about the level of reliability of self-driving cars much much earlier than the hundreds of millions of miles needed for building confidence by focusing just on the fatality rate.

One could object that the relationship between the different types of failure rates is not known. Self-driving cars will certainly have a different correlation between their various failure rates than human drivers. But it is very hard to find plausible arguments why self-driving cars might have a property damage or personal injury accident rate much lower than human drivers but could at the same time have a fatality rate that is significantly higher than human drivers.

Therefore there is no reason why a comparison of the reliability of self-driving cars and human drivers should primarily focus on the fatality rate. We should be glad that fatal accidents are not that frequent which immediately means that enormous numbers of miles need to be driven by self-driving cars if we could only use this measure to estimate the safety of self-driving cars. But many other statistical measures (crash rate, personal injury rate, accident rate etc. are available and much easier to measure. Before we look into these measures in more detail further below, we need to point out additional shortcomings of the current statistical argument:

7.3 Choosing the right independent variable

The number of vehicle miles traveled may not be the best unit against which to measure accidents. The quickest way to accumulate test miles is to drive on highways. But highway driving is safer than driving on many other streets. Of the 3.148 billion vehicle miles traveled in the US in 2015 about a quarter occurred on rural or urban interstate highways; however only about 12% of fatal accidents occurred on these roads. Because the average speed varies between different road types, driving time may be a better basis for measuring failure rate than driving distance. It is well known that urban driving is the most difficult scenario for self-driving cars (and probably also for human drivers, although this may be reflected more in the crash statistics than in the fatality statistics because average speeds are lower). Because average speeds differ greatly, using time instead of distance will have a significant effect on the reliability rates and might significantly reduce the number of test miles (while also reducing the incentive for self-driving car companies to accumulate test miles on what many regards as the "easiest" roads - interstate highways!).

Conceptually, driving is most difficult, when the environment changes, when actions need to be taken. Just

driving straight is not much of a challenge. We should look for other measures that come closer to the complexity of driving and the risks associated with driving. A simple proxy that would be easy to calculate for self-driving cars and not impossible to estimate for human driving statistics could be the number of intersections which are passed. This measure contains both an element of distance and risk because a large percentage of accidents occur at intersections. Therefore the number of intersections passed might also be a good independent variable against which to measure reliability rates.

Just using overall miles driven and not differentiating between the different types of environments which the cars are driven in and not differentiating between the risk levels of the routes is intellectually lazy. Over the years we have collected enormous amounts of information about traffic accidents and the situations in which they occur; we need to carefully go through this data and develop much more refined measures for understanding and assessing the reliability of human as well as autonomous driving.

7.4 Averaging human driving capabilities?

Which brings us to the next problem of our simple statistical model: the reference which we use to determine whether a self-driving car is safe. We have abundant statistics about the fatality and crash rates of human drivers but are these useful measures? Do we really want to include all the accidents caused by drunk drivers, tired drivers, texting drivers, recklessly speeding drivers in this comparison? Can such accidents even remotely be included in the standard against which we want to measure the reliability of self-driving cars? Who would be we willing to tolerate similar levels of errors caused by algorithms? If we think this through, it should become obvious that no society would tolerate self-driving cars that have levels of accidents similar to human drivers. These cars - that never tire, never drink, never take their attention off the road - would have to make other grave errors at much higher rates than humans which our societies would never be prepared to forgive. Although the big averages of human-driven fatality and crash rates are easy to obtain, they are not the standard against which we should measure the reliability of self-driving cars. Instead we will need to carefully craft our expectations and requirements concerning the level of safety which self-driving cars should exhibit into a set of measures based on all that we know about the problem and risk of driving in general, not just of human driving. A real effort is required to build such measures and statistics and the enormous body of research related to safety can be a basis.

7.5 Much more detailed models of driving risk needed

We need to go far beyond the established human-centric look at the problem of driving. We need scientists who build detailed risk models of driving which leverage our accidents statistics and determine the contribution of road structure, congestion levels, weather and many other factors on the risk of driving which can then be used to build reference metrics for the behavior of self-driving cars. We should not let the companies building self-driving cars create these measures all on their own but we need their input because as they develop and test their cars they also gain many deep insights into the structure and risks of the driving problem.

It is a major oversight of the Google team and other commercial developers of self-driving cars that they have not made an effort to show the world better models and metrics for evaluating the safety of driving. Google's monthly reports must be lauded as a level of transparency that is important and that no other developers of self-driving cars (Daimler, Ford, Volkswagen, Nissan, Volvo, Tesla, EasyMile, Uber, Zoox, Baidu, Delphi, and many others) have provided yet. But Google should use their monthly reports not just to provide a mileage count (a number that all readers can misinterpret too easily) but also to educate the readers that there are better ways of measuring the safety levels of a self-driving car. Without giving too much away they could, for example, break their mileage numbers down by road type (at least: city, highway (both interstate and rural), rural). This would help turn the focus of the discussion on the safety of self-driving cars away from a single, overall number. They could also publish a complexity metric which indicates the average complexity of the environment their vehicles have driven through or the overall distribution of complexity encountered during their test drives. Just publishing the number of intersections passed (see above) would also help change the nature of the discussions on assessing the safety of self-driving cars. The public should expect that all companies providing self-driving cars publish metrics about the safety of their cars during testing and that these metrics should go significantly beyond just distance driven. The emerging self-driving car industry should collaborate on putting together such metrics (rather than wait for the regulators to ask for them).

7.6 Reliability of self-driving cars is not a constant

Another key problem of the simple statistical argument needs to be raised: unlike the usual failure rate scenarios, the reliability rate of a self-driving car model can not be regarded as a constant. It changes, improves, over time; and it has some very unusual properties: Self-driving cars can be placed on the road in such a way that a software defect present in all instances of a car model can - upon detection in one instance - be removed from all instances. Unlike in the manufacture of goods where batches are tested to detect design and manufacturing problems, it is possible to remove the defect not just in batches going through the manufacturing process in the future but also in those batches that have already been released to the client! This is a **fundamental difference!** Let us assume that the developer of a self-driving car model releases 1000 cars in a well-managed fleet of self-driving cars and operates them on public roads. The developer has tested the model strenuously over several hundred thousand miles and has performed additional staged tests on test-tracks and run the cars through millions of miles of traffic in a simulator. The fleet quickly accu-

mulates miles and a crash occurs. This changes the estimate of the failure rate. If this new estimate does not exceed the required minimum failure rate for self-driving cars, then the cars can remain in service. The developers will improve their algorithms to handle the problem situation and deploy the fix to all cars on the road: The car model's failure rate (which is not directly observable) is no longer the same. The simple statistical model does not take into account that there is a feedback loop in the failure rate which tends to lower the failure rate with every failure that is encountered.

On the other hand, when a crash occurs, this could also raise the estimated failure rate above the minimum failure rate and indicate that the car model is not suitable for public use. All deployed cars could then be grounded immediately, thus instantly preventing further damage. This is very different from other physical products, such as a medicine or a bicycle where any flaw detected after releasing such products to the consumer will occur again and again because recalls of conventional faulty products are very difficult. This does not mean that we should relax the criteria for deploying self-driving cars. However we need to recognize that we can not simply transfer established reasoning about product failure modes and their consequences to the domain of self-driving cars; the standard statistical modeling approach to this problem is not quite appropriate.

We need to develop a more refined model for estimating the safety of self-driving cars where we account for the unique properties of the problem: there are different types of incidents which indicate failures of different magnitude; we know that the reliability of these cars increases somewhat predictably over time, that models with evidence of too high a failure rate can be grounded immediately, and we know that cars driven by humans also produce accidents.

Most of those worried about the safety of self-driving cars tend to overlook that the problem is dual sided: We need to avoid two cases:

a) That we consider self-driving cars safe and release them to the public although in reality they are not safe enough yet and

b) that we consider self-driving cars unsafe and don't release them to the public although they are safe.

In case a) self-driving cars will cause damage because people relying on their safety will use them. In case b) damage will be caused because people who would have used a self-driving car are now driving themselves.

Thus the problem of determining the reliability of self-driving car safety is more complex than the usual problems where statistical failure rate estimation is applied. More variables and the dynamic dual-sided nature of the problem need to be taken into account. It is time that more elaborate models are developed!

7.7 Intuitions about human driving capability

Before we close, we need to point out two more issues: Our intuitions concerning the safety of self-driving cars may be clouded by the assumptions that we understand the problem of driving well and that overall we are good drivers. This may not be the case as the following thought example shows:

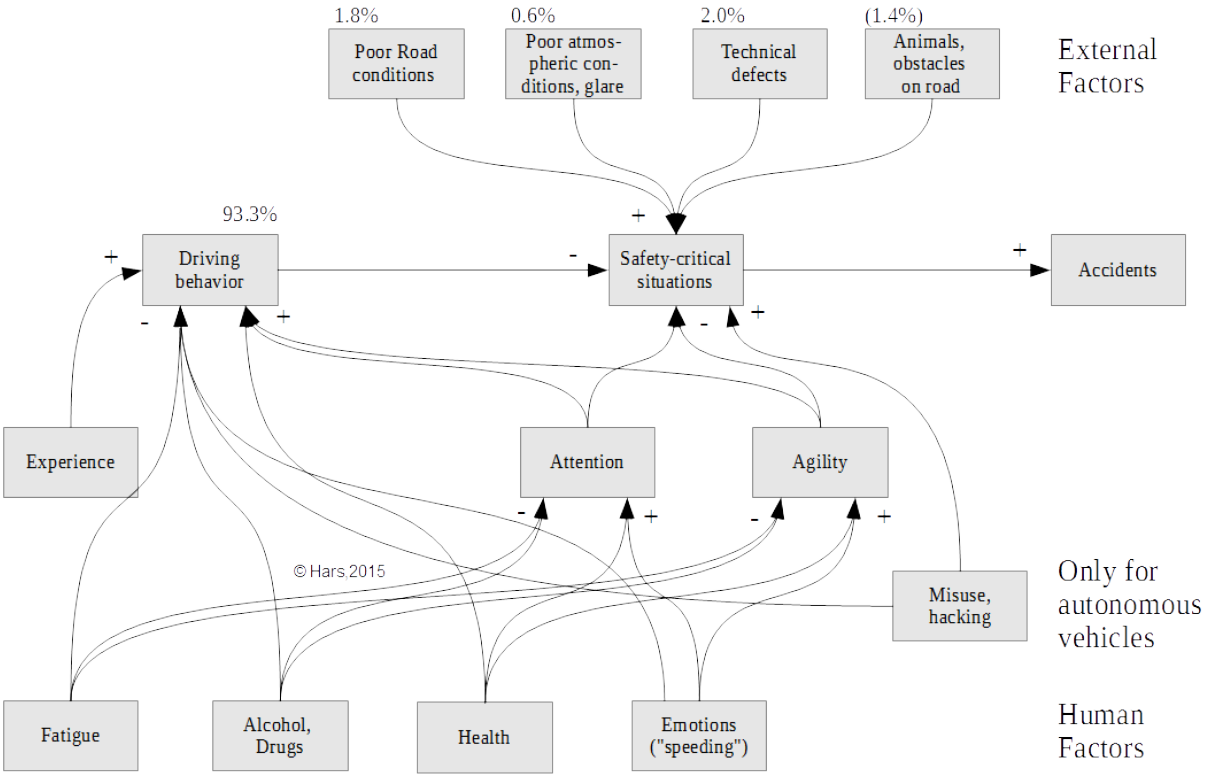
Assume that we have a set of human drivers who all have exactly the same driving capability and share one trait: they randomly take their attention off the road in front of the car for a brief interval of just half of a second every 10 minutes (which translates to a probability of 0.5 / 10 * 60 = 0.083%). Then assume that a certain potentially difficult problem arises directly in front of the cars in a half-second interval with a probability of once per hour (=0.014%). For example, a car in front unexpectedly steps on the brakes. In mst of the cases, the driver reacts immediately, masters the safety-critical situation and nothing happens. But some drivers will be caught in exactly that short interval where they are not paying attention. They can't react in time and an accident results. The probability that this happens is quite low (0.083% * 0.014% or about 1 in 8 million, or once every 1100 hours of driving) but it will happen from time to time. Although we know that all of these drivers drive equally well, the small minority of drivers with an accident will be regarded as poor drivers while all the others will be proud of their spotless driving record. They don't realize that they were just lucky!

Chance (or luck) may play a larger role in human driving than we like to think. That more accidents don't happen might not be so much due to the great capabilities of human drivers but to the relatively low frequency of difficult, unexpected situations. For example, if a car is stopped in the right lane of a highway (lights not flashing), it may not take long - even with good visibility - until someone barrels into it. Fortunately such situations occur rarely and - knowing the danger - people will make an effort to secure the site when such situations occur.

7.8 Safety-critical situations, not just accidents

Thus we need much more refined models of driving risk. Accidents form the basis for our official statistics but accidents are only the tip of the iceberg, as the graphic below shows. Accidents are the outcome of safety-critical situations which have gone wrong. The graphic shows a simple model where driving behavior combined with external factors such as poor road conditions, technical defects, obstacles etc. lead to safety-critical situations. Fortunately, most of these safety-critical situations don't lead to accidents. The percentages next to some of the factors are from the National Crash Causation Survey and relate to the relative frequency that this factor is identified as first factor leading to an accident.

The graphic shows that with more than 93%, driving behavior is a leading factor for fatal traffic accidents. But it also highlights a major difference between human drivers and self-driving cars. With human drivers the focus is on accidents. With self-driving cars, the focus shifts to safety critical situations. Any minute mistake of a self-driving car is cataloged and evaluated, long before any accident happens. When a human driver slightly swerves out of the lane, nobody is concerned (unless an accident happens); however, when a self-driving car does not properly stay in



its lane, then this is - rightly - treated as a defect. The same happens when a self-driving car does not give the right of way to other cars etc. Thus we should not just use accident statistics to compare the reliability of human drivers and self-driving cars, we should compare their driving behavior on the frequency of entering safety-critical situations. This is easy to measure for self-driving cars but requires more effort on the side of human driving. Fortunately with the advances in self-driving car technology we now have the means to collect and analyze large amounts of video feeds of actual human driving and determinew safety critical situations and driver behavior.

7.9 Conclusion

In summary, this chapter shows that the idea that self-driving cars need to drive hundreds of millions of miles before we can be convinced that they are safe is full of flaws. It is misleading to just focus on fatality rates where many other correlated measures for reliability are available that are easier to measure. It is wrong to focus primarily on accidents; the focus should rather be placed on the avoidance of safety-critical situations. Our intuitions related to our own capabilities as a driver may be wrong. Luck may be more important than we like to admit to ourselves. We also need to realize that the reliability of a self-driving car model changes over time and that the damage potential of self-driving cars found to be defective after release to the public is very different from and much smaller than conventional defective products which are hard to recall. Self-driving cars can be effectively and instantly grounded!

Finally, anybody concerned with the safety of self-driving cars needs to consider the dual sided nature of the problem. Not only do we need to avoid self-driving cars being released to the public too early. Because the alternative (human driving) is not safe at all, there is a real risk to releasing self-driving cars too late to the public which will also lead to many traffic deaths that could have been avoided. Almost none of the regulators and safety advocates acknowledge this admittedly very difficult aspect of the regulatory decision problem at all.

It is time to change the discourse on the reliability of self-driving cars and to collaborate to build better, much more refined models of driving and driving risk - both for human drivers and for self-driving cars.

About this document

This is a living document. It will be updated and extended. First version published online on 2015-06-24. Last modified on: 2016-09-30.

Revisions:

2015-07-01: Added Misconception 5: Public demonstrations of self-driving cars provide an indication of their capabilities and added Misconception 6: Self-driving cars need to make ethical judgments

2016-09-30: Added Misconception 7: To convince us that they are safe, self-driving cars must drive hundreds of millions of miles.

Get these and more articles at:

http://www.inventivio.com/publications.html