



## Self-Driving Cars and Engineering Ethics: The Need for a System Level Analysis

Jason Borenstein<sup>1</sup>  · Joseph R. Herkert<sup>2</sup> · Keith W. Miller<sup>3</sup>

Received: 24 August 2017 / Accepted: 6 November 2017  
© Springer Science+Business Media B.V., part of Springer Nature 2017

**Abstract** The literature on self-driving cars and ethics continues to grow. Yet much of it focuses on ethical complexities emerging from an individual vehicle. That is an important but insufficient step towards determining how the technology will impact human lives and society more generally. What must complement ongoing discussions is a broader, system level of analysis that engages with the interactions and effects that these cars will have on one another and on the socio-technical systems in which they are embedded. To bring the conversation of self-driving cars to the system level, we make use of two traffic scenarios which highlight some of the complexities that designers, policymakers, and others should consider related to the technology. We then describe three approaches that could be used to address such complexities and their associated shortcomings. We conclude by bringing attention to the “Moral Responsibility for Computing Artifacts: The Rules”, a framework that can provide insight into how to approach ethical issues related to self-driving cars.

**Keywords** Artificial intelligence · Automation · Engineering ethics · Self-driving cars · Socio-technical systems

---

✉ Jason Borenstein  
borenstein@gatech.edu

Joseph R. Herkert  
jherkert@ncsu.edu

Keith W. Miller  
millerkei@umsl.edu

<sup>1</sup> School of Public Policy and Office of Graduate Studies, Georgia Institute of Technology, 685 Cherry Street, Atlanta, GA 30332-0345, USA

<sup>2</sup> Interdisciplinary Studies, Genetic Engineering and Society Center, North Carolina State University, Raleigh, NC, USA

<sup>3</sup> College of Education, University of Missouri – St. Louis, St. Louis, MO, USA

The current literature on self-driving cars tends to focus on ethical complexities related to an individual vehicle, including “trolley-type” scenarios. This is a necessary but insufficient step towards determining how the technology will impact human lives and society more generally. Ethical, legal, and policy deliberations about self-driving cars need to incorporate a broader, system level of analysis, including the interactions and effects that these cars will have on one another and on the socio-technical systems in which they are embedded.<sup>1</sup> Of course, there are many types of self-driving vehicles that are not cars, including autonomous trucks (Anderson 2015), and they carry with them their own interesting ethical issues. For example, self-driving public transportation, like taxis, could have environmental and other important benefits (Greenblatt and Saxena 2015). Though much of the discussion is applicable to a range of ground-based vehicles, the focus here is on privately owned cars.

Assuming that self-driving cars with varying levels of automation will be operating on the same roads as other vehicles, multiple trolley-type situations (and other complex ethical problems) will emerge. This will require the formation of nearly instantaneous and coordinated decisions by cars, groups of cars, and other entities. The policy challenges emerging from these complexities are starting to be recognized (e.g., see Fagnant and Kockelman 2015). The ethical ramifications of such policy considerations are the focus in this article.

To bring the conversation of self-driving cars to the system level, two traffic scenarios will be discussed; the first involves two types of self-driving cars interacting with one another while sharing the same road. The second scenario contains additional complexity; it involves pedestrians, motorcycles, conventional cars, and animals sharing the same road. These scenarios are intended to illustrate some of the key system level issues that need resolution. For example, if vehicle-to-vehicle communication becomes the norm, will there be sufficient standardization in the design of different makes for the cars to communicate effectively with one another? Will drivers in hybrid human/autonomous vehicles be permitted to out-manuever fully autonomous vehicles by switching in and out of autonomous mode? Possible responses to such complexities include the development of safety and interoperability standards, sophisticated vehicle-to-vehicle communication systems, and technologies for centralized system control (e.g., centralized intersection management). Each of these responses, however, raises a new set of system complexities, regulatory needs, and ethical issues. And it should be kept in mind that our hypothetical scenarios only take into account a relatively finite number of vehicles; in the future, engineers, city planners, and others may have to predict and manage the behavior of hundreds, if not thousands, of self-driving vehicles along with any other automated technologies (such as light rail) or entities that may interact with those vehicles.

This inquiry is motivated by the concern that the rush to bring self-driving vehicles (a complex and rapidly developing technology) to market may compromise consumer safety and autonomy. Laws and regulations for these vehicles must be guided much more consistently and thoroughly by sophisticated ethical analyses of

---

<sup>1</sup> For a detailed discussion of human factors in socio-technical systems see Carayon (2006).

the relevant socio-technical systems. To inform this process, engineers and others should consider the “Moral Responsibility for Computing Artifacts: The Rules”, a framework developed by an ad-hoc interdisciplinary group of computing professionals, engineers, and ethicists (Ad Hoc Committee 2010; Miller 2011). The Rules provide a framework for illustrating some of the key issues that must be addressed.

## Background

A broad range of companies is seeking to gain a foothold in the self-driving car arena. At the time of this writing, at least 44 main players are in this space (CB Insights 2017). Many of them are high-tech enterprises that are designing cars for the first time. The publicly-expressed motivation for creating the technology is to improve road safety. Advocates typically present statistics on highway deaths (ASIRT 2017) and how human drivers are a significant cause of such accidents (NHTSA 2015), and then seek to make the case of how many lives self-driving cars could save (e.g., Setyon 2016).

Economic incentives are obviously a key reason why an eclectic range of companies has such a keen interest in developing the technology. Beyond increased safety, some of the technology’s alleged benefits include lessening the need for parking spaces (Sisson 2016) and expanding the population of people who could make use of cars, such as those who have physical impairments that would make it difficult or impossible to operate current vehicles (Claypool et al. 2017). Self-driving cars may also ease traffic congestion; but as with the other purported benefits, scholarly debates persist about the degree to which, if at all, these benefits would be realized (e.g., Tech Policy Lab 2017).

To assess the technology’s potential benefits and drawbacks, it is important to be specific about which particular types of self-driving cars are being discussed. With that in mind, SAE International (2016) delineates levels of driving automation ranging from 0 to 5 (see Table 1). Level 0 entails that the human driver is in full

**Table 1** Levels of driving automation. (modified from SAE International 2016)

<i>Human driver monitors driving environment</i>	
Level 0—no automation	Full-time operation by human driver
Level 1—driver assistance	Single driver assistance system (steering or acceleration/deceleration)
Level 2—partial automation	Driver assistance systems for both steering and acceleration/deceleration
<i>Automated system monitors driving environment</i>	
Level 3—conditional automation	Automated operation with human driver expected to respond to request for intervention
Level 4—high automation	Automated operation even if human driver fails to appropriately respond to request for intervention
Level 5—full automation	Full-time automated driving system

control of the vehicle's operation and the vehicle does not possess any significant automated features. Levels 1–2 refer to the car having some automated features such as lane correction or parallel parking assistance, but the human driver is not supposed to ever fully relinquish control over the vehicle. A Level 3 system can handle certain driving conditions on its own but a human driver must supervise its operation; the human driver is supposed to stay cognitively engaged in the vehicle's operation in order to take over if a safety-critical situation emerges. Level 4 can function on its own, without human supervision, in a subset of, but not all, driving conditions. Level 5 refers to an automated system that is essentially able to operate the vehicle in the same conditions that a human driver would; in principle, Level 5 automation is designed to handle all safety-critical situations. The subsequent discussion centers primarily on self-driving cars operating at Levels 3–5.

### Necessary but not Sufficient Steps

Ethical analysis of the behavior of a single self-driving car and how it may interact with individual people and other technologies is certainly needed (Borenstein et al. 2017). For example, examining how the technology might respond to a “no win” situation or a true ethical dilemma, such as the Trolley problem, is a vital task (e.g., JafariNaimi 2017; Gogoll and Müller 2017; Nyholm and Smids 2016; Lin 2013). Some scholars are seeking to evaluate empirically how drivers might handle a Trolley-type of situation (Frison et al. 2016); others are considering whether ethical theory, such as the application of Utilitarian reasoning (Bonnefon et al. 2016), may help resolve what is an appropriate response by the car in a crash situation.

Whether a self-driving car will be sophisticated enough to anticipate a pedestrian's behavior is another issue that warrants exploration (Lafrance 2016). What is also a vital part of the conversation is who should be held ethically or legally responsible for an accident involving a self-driving car (Hevelke and Nida-Rümelin 2015). Given that it does not make sense to hold the technology itself responsible for a crash (at least at the present level of sophistication for AI), the courts or other relevant stakeholders must determine the circumstances under which it may be the fault of the designer, manufacturer, car dealer, those riding within the car, and/or some other entity. However, analyses of issues at the individual vehicle level are not sufficient on their own to inform whether it is ethically appropriate to use self-driving cars on public roads. Viewing self-driving cars from the system level helps to call attention to many multi-faceted ethical concerns, including the kinds of socio-political ripple effects embracing the technology might have on society (Blyth et al. 2016).

### Moving to the System Level

The phrase “system level” is being used in two ways. In the first instance, “system level” refers to two or more self-driving cars interacting with one another, other types of vehicles, pedestrians, or animals on the roadway along with any relevant

environmental factors. In a broader context, “system level” refers to the entire enterprise of self-driving car industry and how it interacts with other socio-technical systems, including the networks of manufacturers, regulators, law enforcement personnel, consumers, and the general public. The aim here is to bring attention to ethical issues on a more abstract and broader scale while not ignoring aspects that involve a single car. Many ethical theories and ethics scholars encourage individuals to look beyond themselves, and local concerns, when thinking about justice, values, and morality (e.g., see Bok 1990). However, the ethical approach advocated in this paper was inspired at least as much by the concepts of system engineering as it was by any particular ethical theory.

To illustrate what should be added to the conversation, a description of two driving scenarios is provided. The scenarios highlight ethical and technical complexities that emerge when self-driving cars are viewed from the system level. The hope is that a more nuanced and thorough analysis of the potential effects of the self-driving car will take place as companies are starting to design and deploy the technology, and as regulatory agencies render decisions about the conditions under which the use of the technology is permissible. Note that “traffic” is considered an obstacle when the focus is on a single car; when “traffic” itself becomes a focus in system level thinking, there are both individual aspects of each car and characteristics that belong most properly to the traffic as a group phenomenon.

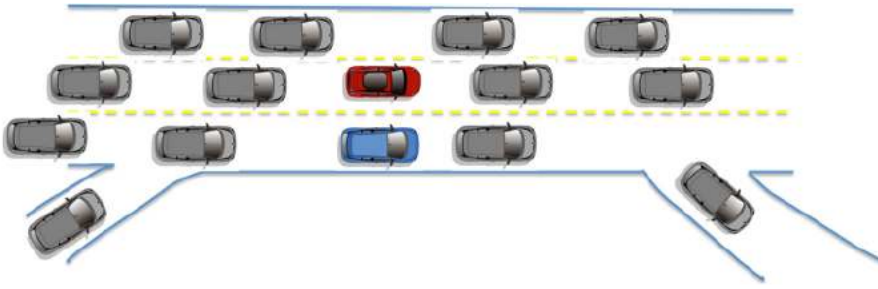
### Scenario One

Suppose that the same road is being used simultaneously by a Level 2–3 car, such as the Tesla S (which for simplicity sake will be referred to as “Car A”) and by a Level 3–4 car, such as the Google spin-off Waymo (“Car B”). A range of circumstances illustrating the interactions between the two models of car could potentially be brought forward; yet this situation involves highway driving. In this first scenario, Car A is currently on a highway and is attempting to merge onto the off-ramp so it can depart the highway at the next exit. Car B is starting to travel on the on-ramp and is attempting to merge onto the highway where Car A resides. At the instant of time on which the scenario focuses, the cars are exactly parallel to one another. To add more context, one should imagine that the traffic is heavily congested, with vehicles both in front of and behind each of the two cars (see Fig. 1 for a diagram illustrating the scenario).

### *Resulting Complexities*

Although not an exhaustive list, some of the key issues that need resolution if the two different models of self-driving cars will operate on the same roads include:

- Is the default expectation that these cars will be able to communicate with one another?
- What will the mechanism be for “encouraging” rival companies (e.g., Tesla and Google) to coordinate design and testing efforts?



**Fig. 1** Car A is red, and hoping to exit. Car B is blue, hoping to merge left. Note that the diagram is not drawn to scale. (Color figure online)

- Will drivers of certain types of self-driving cars require special training and/or licensing (e.g., to mitigate or prevent erratic behavior on the part of human passengers/drivers)?

To avoid a deadlock situation between the two cars competing for position, a possible strategy would be to give one of the cars priority over the other. If this is a reasonable approach, what will the protocol be for establishing this priority? Will all cars, at least out of those communicating with one another, be obliged to accept the decision of which car gets priority? And who or what will make and enforce this decision? Moreover, will there be an attempt to enforce the decision on cars driven by humans as well as on driverless cars?

The situation becomes more convoluted because at least some self-driving cars are designed (e.g., Level 3 automation) so that a human driver can, and at times is intentionally supposed to, take over their operation. Thus, what is the default expectation going to be in potentially dangerous situations? At the present time, for cars such as the Tesla S, the design entails that the human driver is expected to respond (Lee 2017), but that presupposes that the driver is paying close enough attention to the road, recognizes the need for intervention, has adequate time to react, and knows how to do so. If the default setting changes and the automated system is supposed to respond, then the human driver will need to be aware of this design pathway. Otherwise the human driver could interfere with the performance of a safety critical feature. In short, having a human in the loop adds much complexity in the effort to generate a reliable prediction of how the car will behave (Casner et al. 2016); yet, on the other hand, given the limits of AI (Çelikkanat et al. 2016) and the interactive complexity and tight coupling of self-driving vehicles, a human operator is potentially needed to handle chaotic or difficult to anticipate situations (Perrow 1999).

Proponents of self-driving vehicles might reply to this scenario by suggesting that it is unrealistic because the Waymo, or another sophisticated self-driving car, would “know” how to adjust its location and speed. They may argue, for example, that Light Detection and Ranging (LIDAR) sensors or other devices will enable the car to adjust its speed accordingly so that it would not be completely parallel with another car when attempting to merge. Yet if the traffic is congested enough where

human drivers in other cars on the road are tailgating or taking part in other less than ideal driving behaviors, this type of scenario might not be fully avoidable.

Furthermore, the two self-driving cars may not respond correctly to one another due to design differences and compounding variables such as the unpredictability of the driver in Car A. The driver might try to intervene at any given time, or due to inattentiveness or other reasons, may decide not to take over the car's operation. A reasonable case could be made that this kind of situation is bound to occur unless both cars are completely automated, which may include external control of both cars by a third agent. Even in that circumstance, however, complex interactions and tight coupling could occur among the two cars and the controller agent, resulting in a variety of unanticipated safety concerns (Perrow 1999).

## Scenario Two

To add further layers of complexity, suppose a busy intersection or another portion of a city road is simultaneously being used by Car A and Car B along with pedestrians, a motorcyclist, a bicyclist, a person driving a regular (Level 0) automobile while texting, and a squirrel. Assuming that the two self-driving cars can effectively and quickly communicate with one another (a non-trivial assumption), the interactions between and among these different entities makes what has to be anticipated much more convoluted and messy.

### *Resulting Complexities*

If the two different models of self-driving car are going to be allowed on the same road, like what is described in Scenario Two, some of the key issues that need resolution include:

- Will human drivers, bicyclists, and pedestrians try to exploit the safety features of a self-driving car (e.g., by weaving in and out of self-driving traffic)?
- What happens if a bicyclist, pedestrian, or other entity attempts to anticipate an autonomous car's behavior (at Level 4 or 5) but is unfamiliar with or misunderstands the capabilities of the car (e.g., by assuming all self-driving cars have human backup)? Or what happens if that entity interacts with a different model of self-driving car than was anticipated?
- Will human drivers experience "road rage" if self-driving cars are strictly following the law? If so, how should the self-driving cars behave in response?

Notice that human drivers have for some time now been making decisions about issues similar to those mentioned above though admittedly they have not always done so effectively, consistently, or safely. The systems guiding driverless vehicles will need to emulate, and perhaps improve upon, the human ability to grapple with complexity, ambiguity, and uncertainty.

It is reasonable to assume that at least some individuals will try to exploit the safety features built into self-driving cars. And if and when multiple vehicle models are on the same roads, the interaction between human beings and technology is

going to be a dynamic and chaotic experiment. For example, a wicked and multifaceted problem for designers of self-driving cars is anticipating the behavior of pedestrians given their fluctuating levels of attentiveness and willingness to adhere to traffic laws, and their varying risk tolerances. Conjoin this with other factors such as the weather (e.g., rain or ice), being late for a meeting, and frustration leading to impatience, designers will have a significant engineering challenge. Map onto that additional models of self-driving cars and other vehicles (e.g., bicycles) sharing the road, and a seemingly endless stream of confounding variables emerges.

Designers are working on possible technical fixes, such as having bicycles communicate with self-driving cars (Krauss 2017), that perhaps may help ease this sort of problem. Yet it is unclear whether a technical fix can fully address the issue, including when it comes to unanticipated behavior of animals both domestic and wild. Moreover, that kind of approach greatly increases the amount of data that must be processed almost instantaneously.

## Possible System Level Approaches

In this section, three candidate ideas are outlined that supporters of self-driving cars might try to use to mitigate system level problems with the technology. The first approach involves promulgation of technical standards while the other approaches rely on augmenting self-driving vehicles with additional technologies including communication systems and infrastructure improvements. It is noteworthy that each proposed approach arguably suffers from significant shortcomings. They carry with them, at least for the foreseeable future, a collection of unresolved concerns. While this list of approaches is not exhaustive, the categories selected—regulatory, vehicle design, and infrastructure design—are representative of three primary means of addressing the overall safety of self-driving cars.

### Safety/Interoperability Standards

The US government is starting to take more of an active role in regulating self-driving cars (Rogers 2017); one step in this direction is the creation of a 15-point checklist (USDOT 2016; Kang 2016). The checklist primarily focuses on the individual self-driving car with only a few items giving a nod to the complex interactions among multiple self-driving cars and the sociotechnical systems in which they are embedded (see Table 2). A necessary step to promote the safe operation of the technology is standardization with regard to safety critical operations across different self-driving car companies. Thus, recognizing the need for at least some level of uniformity in the design and performance of self-driving cars, safety and interoperability standards are likely to be a necessity.

Yet the mechanism at this point for encouraging or requiring car manufacturers to standardize safety rules is undetermined. For example, how much space behind another vehicle should the automated system allocate at various speed intervals? Admittedly, having strict uniformity in safety rules when the vehicles have a range



**Table 2** U.S. federal checklist for autonomous vehicles. (categories derived from NHTSA 2016)

	Vehicle focus	System focus
Consumer education and training	X	
Crashworthiness	X	
Cybersecurity	X	
Data recording and sharing	X	
Detection and response	X	X
Ethical considerations	X	
Fall back	X	
Human-machine interface	X	
Laws and practices	X	X
Operational design	X	
Post-crash behavior	X	
Privacy	X	
Registration and certification	X	
System safety	X	
Validation methods	X	

of designs and performance (different weights, braking systems, etc.) may not be feasible. Conjoining this to the status quo where the default norm is companies that are in this space are intensely competitive, as opposed to cooperative, the likelihood of early voluntary standardization seems vanishingly small.

### Vehicle-to-Vehicle Communication (V2V)

For self-driving cars to reach their potential in terms of enhancing safety, some may argue that networks must be established that enable the cars to communicate with each other and potentially other entities (Beuse 2015). A technical approach of vehicle-to-vehicle (V2V) communication may address system level problems; it would entail having all self-driving, and perhaps other, vehicles communicate with one another in order to facilitate coordinated vehicle operation (Harding et al. 2014). Issues that need resolution for a successful implementation of V2V include:

- What information will the system require to be broadcasted by any cars on the road?
- Who will standardize communication protocols?
- How will the reliability of communication be established and certified?
- How will “rules of engagement” be established, and who or what will enforce these rules?
- Will the system force human driven cars to stop or slow down in potentially dangerous situations?

An overarching variable is whether regulatory agencies are going to mandate that all self-driving cars have connectivity capacity. Intellectual property ramifications also emerge from the sharing of information among vehicles (Crane et al. 2017). For

example, will one company gain proprietary information from a rival once such sharing occurs?

Another related concern is whether V2V will increase the vulnerability to hacking. Receiving such information certainly opens up pathways for external parties to exploit the weaknesses in a car's software and hardware. Along these lines, vehicle-to-infrastructure (V2I) communication can raise similar (unresolved) issues about the security and ownership of transmitted data (Crane et al. 2017). According to Lewis et al. (2017), "V2V and V2I communications are still in pilot phases and will require significant standardization and investment to become widespread."

### Centralized Intersection Management

A third possible technical approach to system level issues is centralized control of vehicle operation such as Centralized Intersection Management whereby self-driving vehicles are "dispatched" through busy intersections with the aid of technologies embedded in the intersection infrastructure (Ockedahl 2016). The purported benefits of this type of intersection include increased efficiency and enabling vehicles to travel at a higher speed (Ackerman 2016). But one must consider whether the benefits of such an approach will outweigh its drawbacks. Among the questions posed by Centralized Intersection Management include:

- Will such control systems require complete Level 5 automation to function properly?
- If not, who has priority if there is a mix of Level 3–5 vehicles?
- How will pedestrians and non-automated vehicles (e.g., bicycles) be "dispatched"?
- What happens when the control system is down? Will back-up traffic lights be needed?
- Can the system be effectively protected from cyber and physical attacks?

Intertwined with many of these issues is whether the public will accept complete automation in part because of the resulting loss of autonomy. Many drivers think that having control over their automobile represents an important source of freedom over their own lives (Moor 2016). If the ultimate aim is to reduce the risks emerging from mixing self-driving cars with human-driven cars (and the risks are likely substantial), then the resulting approach may require severe restrictions on human driven cars on most public roads. Being transparent about this eventuality might lead to strong opposition to automated driving in, for example, the U.S., which has a long history of intense interest in personal automobiles. As Moor (2016) suggests, "what happens to the American myth when you take the driver out of it?"

The specter of "Normal Accidents", resulting from interactive complexity and tight coupling of systems (Perrow 1999), must also be taken into account. In other words, since Centralized Intersection Management would require a complex system with various intertwined technologies, it is bound to have "normal" failures. Along these lines, if the system goes down due to weather (ice, flooding, etc.) or another

type of emergency, protocols will need to be in place to handle the traffic flow. When standard traffic lights stop working, drivers in the United States typically treat an intersection like a four-way stop. But the efficiency and safety of that behavior greatly depends on each driver being aware of the “local norm” and actually adhering to it. It is commonplace that drivers violate the norm and try to take advantage of others who are following the rules.

Designers, city planners, and others will have to anticipate what the potential magnitude of harm is if the system controlling the intersection is offline. What would the worst-case scenario look like? Would head-on collisions result? Will effective mechanisms be in place to force some or all vehicles to slow down or stop? Systems engineers and others will need to conduct extensive analyses to determine, for example, to what degree the dispatching process should rely on distributed decision-making in case the centralized system is not functioning properly. Moreover, the manner in which different entities are dispatched could generate ethical problems. Mladenovic and McPherson (2016) claim that traffic control technology, like what is alluded to here, can raise social justice concerns by, for example, assigning a higher priority to a car than a pedestrian.

Furthermore, Centralized Intersection Management will require substantial investment in infrastructure. And it is an open question whether the political will and resources to support it are going to be available. Not only is designing and building such an intersection expensive, but as mentioned above, its implementation may necessitate creating back-up systems as well.

## System Alternatives

A complete system analysis would consider other transportation modes as alternatives to large scale deployment of self-driving cars. A push towards the creation of a fleet of Level 5 self-driving cars, for example, would likely entail substantial infrastructure changes (including “smart” roadways and advanced communication systems (Malesci 2017)) that would come at an enormous economic cost, and pose a wide range of technical, logistical, legal, policy, and ethical challenges. For instance, eminent domain would be a consistent theme from a policy perspective because presumably, large swaths of private land, especially in metropolitan areas, would need to be obtained. Furthermore, city planners and others may need to consider whether it is a more reasonable design pathway to allow these cars to operate on the same roads as other transportation technologies or to isolate the different modes of transportation from one another to the greatest extent possible.

Interestingly, an alternative means for providing ground transportation currently exists in many regions of the world, and it is often already isolated from roads: rail and light rail. Some, but not all, current trains are autonomous vehicles, and they have firmly been embedded in a sociotechnical system for quite some time. The technology has been shown to be able to move large volumes of people in relative safety. For example, the Shinkansen, a Japanese high-speed train, has been in use since 1964, and zero reported accidental fatalities have been connected to its

operation (Economist 2014). In addition, “[A]utonomous public transport offers great potential for the development and promotion of sustainable mobility concepts” (Pakusch and Bossauer 2017). Thus, it becomes relevant to ask whether improving on and expanding existing rail and light rail technology might be a more practical, and more effective, solution to improving transportation than encouraging or mandating wide-spread use of self-driving cars. Such considerations are not something dwelt on in this paper, but focusing on the system level could be a prelude to questioning the entire enterprise of the self-driving car.

## Moral Responsibility for Computing Artifacts: The Rules

The Moral Responsibility for Computing Artifacts is a collection of five rules, championed by Keith Miller (2011) and other computer scientists, engineers, and ethicists (Ad Hoc Committee 2010). These Rules were crafted to provide guidance to the computing and engineering communities especially with respect to pervasive and autonomous technologies. The focus here is on Rules 1 and 4 because they can directly and saliently draw attention to the responsibilities of designers and others who will be involved in the creation and deployment of self-driving cars.

### Rule 1

Rule 1 states that:

The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, and for the foreseeable effects of that artifact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artifact as part of a sociotechnical system (Ad Hoc Committee 2010).

Applying this Rule to this context, the argument can be made that engineers and others have a shared responsibility for the ethical design, development, and deployment of self-driving cars. Admittedly, precisely articulating the scope of “foreseeable use” of self-driving cars by drivers, passengers, and others is problematic. For example, the complexities outlined above greatly complicate any reasonable testing strategy for a transportation system that includes driverless cars. The potential interactions, including hacking and software failures, geometrically increase the number of possible scenarios that should be explored in a responsible testing regime.

How human occupants will behave in Level 3 vehicles, for instance, is largely an open question. While expectations of designers may be “clear”, the actual response of human occupants could range from complacency to over-reaction. Even less clear is the behavior of non-occupants such as pedestrians, drivers of non-automated vehicles, and animals. Nonetheless, designers, users, and others need to make a rigorous and thorough attempt to try to anticipate future uses (and misuses) of the technology.

Despite the challenges inherent in assigning responsibility in this situation, the Rule is still a practical way to encourage accountability in this arena. It is not defensible to abandon the idea of moral responsibility for self-driving cars because of the difficulty of determining (especially *ex post facto*) who was responsible for a particular action or inaction that caused a harm. Instead, engineers should embrace responsibility as something inherent in the process of design and development. Clearly, developers are in a privileged position to know the details of implementation, and this will help them make informed (though clearly not omniscient) predictions about the behavior of a system of driverless cars. Making a sincere and diligent effort to foresee potential consequences of one's work is integrally linked to what it means to be an ethical professional.

At the time of this writing, the authors are not aware of any laws or regulations that establish a minimal standard of testing of Level 3–5 software or hardware. The entities designing, developing, and deploying these vehicles are largely self-regulating. Even if all these entities are making sufficiently good faith efforts to conduct testing (which seems overly optimistic), safety concerns in this rapidly changing environment may require some kind of organized initiative to ensure quality control that transcends voluntary, ad-hoc measures. Moor (1985) points out that rapidly developing technologies often result in “policy vacuums” in which previous laws, regulations, and customs are inappropriate for the new artifacts and systems. Arguably, self-driving cars are being developed and operate in such a policy vacuum; this is an area that could use system level, societal attention in the immediate future.

#### Rule 4

Rule 4 states that:

People who knowingly design, develop, deploy, or use a computing artifact can do so responsibly only when they make a reasonable effort to take into account the sociotechnical systems in which the artifact is embedded (Ad Hoc Committee 2010).

Engineers and others must consider the sociotechnical systems in which self-driving cars are embedded when they are designing and testing the technology. This includes evaluating how safety features may be impacted by interactions among drivers, passengers, pedestrians, all types of vehicles, infrastructure, and the external environment. For example, how will self-driving vehicles react to unusual traffic patterns such as accident or construction backups, funeral processions, sporting or other entertainment events, or natural disasters? Moreover, will they be sophisticated enough to detect what a person on a sidewalk who waves a hand is trying to convey? The automated system would need to distinguish between someone hailing a taxi versus a police officer attempting to get cars to stop moving (Eustice 2015). In short, self-driving vehicles may need to have the capability of picking up on a range of verbal, visual, and other cues that humans rely on to communicate information; this design challenge is of course complicated by the fact that these cues are not always used in a consistent and clear manner.

Admittedly, the moral requirement conveyed by Rule 4 is not straightforward. It is a daunting task to take into account the vast sociotechnical systems that will be affected (directly and indirectly) by driverless cars. But unless the people who develop and promote these technologies are willing to make honest and sincere efforts to take those systems into account, they are ignoring a profound moral duty. In practice, following this Rule will require a shift in perspective, a shift that will cost time and money. Experts in fields as wide-ranging as civil engineering, urban planning, sociology, psychology, law, and policy will need to consult with the engineers and scientists planning and developing driverless cars. Arguably, this kind of effort should have taken place prior to allowing self-driving cars on public roads.

## Conclusion

The self-driving car enterprise is rapidly moving forward with little or no attention to alternative modes of transportation. While the authors advocate further consideration of alternatives, such as light rail, it is recognized that the momentum behind the autonomous private passenger vehicle is proceeding almost inexorably. Thus, ethical evaluations of self-driving cars must include a *system level* analysis of the interaction that these cars will have with one another, and with the broader socio-technical systems within which they are embedded. Standardization efforts to date have focused almost exclusively at the individual vehicle level. And the law is struggling to keep up when the focus has only been at that scale (Greenblatt 2016). Many technical and ethical questions need to be answered before technologies such as V2V and Centralized Intersection Management become a reality. Because some types of autonomous vehicles are already on public roads, laws and regulations mandating more rigorous testing, including testing aimed at resolving system level issues, is overdue.

## References

- Ackerman, E. (2016). The scary efficiency of autonomous intersections. *IEEE spectrum*, March 21. <http://spectrum.ieee.org/cars-that-think/transportation/self-driving/the-scary-efficiency-of-autonomous-intersections>. Accessed August 21, 2017.
- Ad Hoc Committee for Responsible Computing. (2010). Moral responsibility for computing artifacts: Five rules, version 27. <https://edocs.uis.edu/kmill2/www/TheRules/>. Accessed August 19, 2017.
- Anderson, B. (2015). Self-driving trucks could rewrite the rules for transporting freight. *Forbes*, December 8. <https://www.forbes.com/sites/oliverwyman/2015/12/08/self-driving-trucks-could-rewrite-the-rules-for-transporting-freight/>. Accessed August 16, 2017.
- Association for Safe International Road Travel (ASIRT). (2017). Annual global road crash statistics. <http://asirt.org/initiatives/informing-road-users/road-safety-facts/road-crash-statistics>. Accessed June 22, 2017.
- Beuse, N. (2015). Statement of nathaniel beuse, associate administrator for vehicle safety research, national highway traffic safety administration. Before the house committee on oversight and government reform hearing on “The Internet of Cars”, November 18. <https://www.transportation.gov/content/internet-cars>. Accessed October 19, 2017.

- Blyth, M. N., Mladenović, B. A., Nardi, H. R., Ekbia, N., & Su, M. (2016). Expanding the design horizon for self-driving vehicles: Distributing benefits and burdens. *IEEE Technology and Society Magazine*, 35(3), 44–49.
- Bok, S. (1990). *Common values*. Columbia: University of Missouri Press.
- Bonnefon, J., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576.
- Borenstein, J., Herkert, J., & Miller, K. (2017). Self-driving cars: Ethical responsibilities of design engineers. *IEEE Technology and Society Magazine*, 36(2), 67–75.
- Carayon, P. (2006). Human factors of complex sociotechnical systems. *Applied Ergonomics*, 37(4), 525–535.
- Casner, S. M., Hutchins, E. L., & Norman, D. (2016). The challenges of partially automated driving. *Communications of the ACM*, 59(5), 70–77.
- CB Insights. (2017). 44 corporations working on autonomous vehicles, May 18. <https://www.cbinsights.com/research/autonomous-driverless-vehicles-corporations-list/>. Accessed July 22, 2017.
- Çelikkanat, H., Orhan, G., Pugeault, N., Guerin, F., Şahin, E., & Kalkan, S. (2016). Learning context on a humanoid robot using incremental latent dirichlet allocation. *IEEE Transactions on Cognitive and Developmental Systems*, 8(1), 42–59.
- Claypool, H., Bin-Nun, A., & Gerlach, J. (2017). Self-driving cars: The impact on people with disabilities. Ruderman Family Foundation. [http://secureenergy.org/wp-content/uploads/2017/01/Self-Driving-Cars-The-Impact-on-People-with-Disabilities\\_FINAL.pdf](http://secureenergy.org/wp-content/uploads/2017/01/Self-Driving-Cars-The-Impact-on-People-with-Disabilities_FINAL.pdf). Accessed June 22, 2017.
- Crane, D. A., Logue, K. D., & Pilz, B. C. (2017). A survey of legal issues arising from the deployment of autonomous and connected vehicles. *Michigan Telecommunications and Technology Law Review*, 23, 191–320.
- Eustice, R. (2015). University of Michigan's work toward autonomous cars. [http://www.umtri.umich.edu/sites/default/files/Ryan.Eustice.UM\\_Engineering.IT\\_2015B.pdf](http://www.umtri.umich.edu/sites/default/files/Ryan.Eustice.UM_Engineering.IT_2015B.pdf). Accessed July 5, 2017.
- Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77, 167–181.
- Frison, A-K, Wintersberger, P., & Riener, A. (2016). First person trolley problem: Evaluation of drivers' ethical decisions in a driving simulator. In *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications* (AutomotiveUI '16 Adjunct). ACM, New York, NY, USA, pp. 117–122.
- Gogoll, J., & Müller, J. F. (2017). Autonomous cars: In favor of a mandatory ethics setting. *Science and Engineering Ethics*, 23(3), 681–700.
- Greenblatt, N.A. (2016). Self-driving cars will be ready before our laws are. *IEEE Spectrum*, January 16. <http://spectrum.ieee.org/transportation/advanced-cars/selfdriving-cars-will-be-ready-before-our-laws-are>. Accessed May 9, 2017.
- Greenblatt, J. B., & Saxena, S. (2015). Autonomous taxis could greatly reduce greenhouse-gas emissions of US light-duty vehicles. *Nature Climate Change*, 5(9), 860–863.
- Harding, J., Powell, G., R., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., Lukuc, M., Simons, J., & Wang, J. (2014). Vehicle-to-vehicle communications: Readiness of V2V technology for application. Report No. DOT HS 812 014. Washington, DC: National Highway Traffic Safety Administration.
- Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics*, 21, 619–630.
- JafariNaimi, N. (2017). Our bodies in the trolley's path, or why self-driving cars must \*not\* be programmed to kill. *Science, Technology, & Human Values*. <https://doi.org/10.1177/0162243917718942>.
- Kang, C. (2016). The 15-point federal checklist for self-driving cars. *The New York Times*, September 21, 82.
- Krauss, M. J. (2017). Bikes may have to talk to self-driving cars for safety's sake. *NPR: All Tech Considered*, July 24. <http://www.npr.org/sections/alltechconsidered/2017/07/24/537746346/bikes-may-have-to-talk-to-self-driving-cars-for-safety-sake>. Accessed August 1, 2017.
- LaFrance, A. (2016). Will pedestrians be able to tell what a driverless car is about to do? *The Atlantic*, August 30. <http://www.theatlantic.com/technology/archive/2016/08/designing-a-driverless-car-with-pedestrians-in-mind/497801/>. Accessed August 16, 2016.
- Lee, T. (2017). Car companies' vision of a gradual transition to self-driving cars has a big problem. *Vox*, July 5, <https://www.vox.com/new-money/2017/7/5/15840860/tesla-waymo-audi-self-driving>. Accessed July 22, 2017.

- Lewis, P., Rogers, G., & Turner, S. (2017). Beyond speculation: Automated vehicles and public policy—An action plan for federal, state, and local policymakers. *Eno Center for Transportation*. [https://www.enotrans.org/wp-content/uploads/2017/04/AV\\_FINAL-1.pdf](https://www.enotrans.org/wp-content/uploads/2017/04/AV_FINAL-1.pdf). Accessed August 1, 2017.
- Lin, P. (2013). The ethics of autonomous cars. *The Atlantic*, October 8. <http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/> Accessed August 1, 2017.
- Malesci, U. (2017). Why highways should isolate self-driving cars in special smart lanes. *Venture Beat*, April 25. <https://venturebeat.com/2017/04/25/why-highways-should-isolate-self-driving-cars-in-special-smart-lanes/>. Accessed August 17, 2017.
- Miller, K. (2011). Moral responsibility for computing artifacts: ‘The Rules’. *IT Professional*, May/June, 57–59. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5779006>. Accessed July 21, 2017.
- Mladenovic, M. N., & McPherson, T. (2016). Engineering social justice into traffic control for self-driving vehicles? *Science and Engineering Ethics*, 22(4), 1131–1149.
- Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, 16(4), 266–275.
- Moor, R. (2016). What happens to American myth when you take the driver out of it? The self-driving car and the future of the self. *New York Magazine*, October 16. <http://nymag.com/selectall/2016/10/is-the-self-driving-car-un-american.html>. Accessed August 16, 2017.
- National Highway Traffic Safety Administration (NHTSA). (2015). Critical reasons for crashes investigated in the national motor vehicle crash causation survey. <http://www-nrd.nhtsa.dot.gov/pubs/812115.pdf>. Accessed June 22, 2017.
- National Highway Traffic Safety Administration (NHTSA). (2016). Federal automated vehicles policy: Accelerating the next revolution in roadway safety. <https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016>. Accessed October 19, 2017.
- Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical Theory and Moral Practice*, 19(5), 1275–1289.
- Ockedahl, C. (2016). Could autonomous vehicles be end of the road for traffic lights? *Forbes.com*, April 18. <http://www.forbes.com/sites/trucksdotcom/2016/04/18/could-autonomous-vehicles-kill-traffic-lights/>. Accessed July 21, 2017.
- Pakusch, C. & Bossauer, P. (2017). User acceptance of fully autonomous public transport. In *Proceedings of the 14th international joint conference on e-business and telecommunications* (Vol. 4: ICE-B, pp. 52–60), Madrid, Spain.
- Perrow, C. (1999). *Normal accidents: Living with high-risk technologies* (2nd ed.). Princeton: Princeton University Press.
- Rogers, G. (2017). House E&C considering 16 automated vehicle bills—Here’s what’s in them. *Eno Transportation Weekly*, June 5. <https://www.enotrans.org/article/house-ec-considering-16-automated-vehicle-bills-heres-whats/>. Accessed July 22, 2017.
- SAE International. (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles J3016. [http://standards.sae.org/j3016\\_201609/](http://standards.sae.org/j3016_201609/). Accessed July 17, 2017.
- Setyon, J. (2016). NTSB chairman: Driverless cars could save 32,000 lives a year. *CNSnews.com*, July 1. <http://www.cnsnews.com/news/article/joe-setyon/ntsb-chairman-driverless-cars-could-save-32000-lives-year>. Accessed June 22, 2017.
- Sisson, P. (2016). Why high-tech parking lots for autonomous cars may change urban planning. *Curbed*, August 8. <https://www.curbed.com/2016/8/8/12404658/autonomous-car-future-parking-lot-driverless-urban-planning>. Accessed June 22, 2017.
- Tech Policy Lab. (2017). Driverless Seattle: How cities can plan for automated vehicles. University of Washington. [http://techpolicylab.org/wp-content/uploads/2017/02/TPL\\_Driverless-Seattle\\_2017.pdf](http://techpolicylab.org/wp-content/uploads/2017/02/TPL_Driverless-Seattle_2017.pdf). Accessed July 7, 2017.
- The Economist. (2014). What a ride, October 4. <http://www.economist.com/news/asia/21621880-they-have-been-rolling-50-years-and-without-fatal-accident-what-ride>. Accessed July 22, 2017.
- US Department of Transportation (USDOT). (2016). Automated vehicles policy fact sheet overview. <https://www.transportation.gov/AV-factsheet>. Accessed January 17, 2017.