ISTOCK

Jason Borenstein, Joseph Herkert, and Keith Miller

When you come to a fork in the road, take it.
—Yogi Berra [1]

In the wake of the exposure of Volkswagen's diesel engine test-rigging, a *Bloomberg Business* journalist described the company as "driven by engineering-crazed executives" [2] and *The New York Times* ran a story noting how with today's complex computer systems in automobiles, there are numerous opportunities for misdeeds both by automakers and hackers [3].

Digital Object Identifier 10.1109/MTS.2017.2696600
Date of publication: 8 June 2017

With the advent of so-called autonomous or self-driving cars, such issues may become even more pervasive and problematic. From a legal perspective, a key focal point is who would be at fault if and when an accident occurs [4]. Much also has been written about the ethical complexities posed by self-driving cars [5]–[6]. In accordance with Moore's Law, "(a)s technological revolutions increase their social impact, ethical problems increase" [7]. Yet relatively little has been said about the ethical responsibilities of the designers of self-driving cars.

In this paper, we review Richard De George's classic article on the moral responsibilities of engineers in the infamous Pinto case, and consider whether his analysis is valid in an era of pervasive and autonomous technologies [8]. We undertake a contemporary analysis of the topic as it pertains to engineers who are designers

of self-driving cars, including by applying the "Moral Responsibility for Computing Artifacts: The Rules," a framework developed by an ad hoc interdisciplinary group of computing professionals, engineers, and ethicists [9]. While engineers and engineering managers are not necessarily "crazed," we argue that ethical analysis needs to be integral to the design of self-driving vehicles. Engineering and other relevant communities need to engage with the issue of what it means to uphold one's ethical and professional responsibilities in the era of these vehicles. Designers of the technology should diligently and creatively exercise their moral sensitivity capacities in order to uphold their obligations to the public. Integrating this activity into their decision-making process is a critical element in the realm of "anticipatory technology ethics" as defined by Brey [10] and in the realm of "responsible research and innovation" as described by Sutcliffe [11].

Brey's anticipatory technology ethics requires diligence in terms of forecasting consequences; cross-referencing a technology's description with moral values and principles; and evaluating and elaborating on the ethical issues identified [10]. Sutcliffe contends that responsible research and innovation includes an emphasis on the involvement of society throughout a technology's development; paying close attention to ethical and environmental impacts; effective oversight mechanisms; and openness and transparency. [11] As we shall see, self-driving cars bring all of these aspects into sharp focus.

Scope of Use and Potential Benefits of Self-Driving Cars

Self-driving cars are starting to make their way onto the roads in the United States and in other countries. Many automobile manufacturers, including Mercedes-Benz, General Motors, Toyota, and Tesla have a keen interest in creating these cars. It is interesting to note that Silicon Valley companies, including Google, Apple, and Uber, are seeking to become key players in this market even though their history is not directly tied to the production of cars [12]. According to Edward Taylor, by 2025 projected global sales of "semi-automated" cars is approximately 22.7 million, whereas for "highly automated" cars it is 9 million [13]. At the present time, driving responsibilities within a self-driving car are normally shared between a human being and a computing system. The degree to which this hybrid design cedes control to a human versus an automated system varies greatly depending on the company that created the car.

The touted benefits of such cars include improved safety, in part because they could remedy problems associated with distracted driving and other human driver errors. The U.S. National Highway Traffic Safety Administration suggests the "critical reason" that accidents occur is attributable to the driver approximately 94% of the time [14]. The Association for Safe International Road Travel claims that "(n)early 1.3 million people die in road crashes each year" and "(a)n additional 20–50 million are injured or disabled" due to such crashes [15]. According to computer scientist Moshe Vardi, "So by automating driving, we could save about a million lives a year" [16]. Along related lines, advocates of self-driving cars suggest that the accidents involving them are normally the fault of human drivers and not the technology [17]. Allegedly, these cars could also increase fuel economy, decrease traffic congestion, and may ease parking-related problems [18]. Furthermore, the technology may grant more mobility to those who are currently unable to drive, including those with disabilities [19].

The U.S. government has indicated its intent to support self-driving cars. In January 2016, the Obama Administration announced that in the FY 17 budget, it would request a $4 billion investment over ten years for technology R&D and infrastructure improvements related to self-driving cars [20].

Our Focus

Much of the discussion at the intersection of ethics and self-driving cars has to date focused on high-level ethical dilemmas that might be encountered by a self-driving car, such as the trolley problem [5], [21]–[22]. While these dilemmas are important, many other, subtler ethical issues relating to self-driving cars demand the attention of engineering, engineering ethics, and other related communities. Discussion has also tended to focus on the role of programmers or "coders" in dealing with ethical dilemmas and self-driving cars [23]. While the line between a "designer" and a "coder" is not always sharp, we offer the following distinction:

- Designer — has a say in determining design pathways (e.g., whether the system will rely on user input); responsible for higher level decisions.
- Coder — largely tasked with implementing what the designer specifies; responsible for lower level decisions.

Although not all automotive designers are engineers, engineers are usually involved in the design, development, and testing of safety-critical systems. We seek, then, to discuss the ethical responsibilities of design engineers (hardware and software) throughout the process of the design, development, and testing of self-driving cars.

De George and the Pinto Case

The series of Ford Pintos built in the 1970s is a frequent jumping off point for analyzing engineering decision-making and an engineer's ethical responsibilities. De George provides a thorough examination of the Pinto

case, largely through the lens of what an engineer should be ethically permitted or even required to do in response to a situation of that type (8). More specifically, what should an engineer do if the placement of the Pinto's rear fuel tank might cause harm to the public? While we are not committed to the view espoused in the article, the article is important to examine for at least two reasons. First, De George's article helps to bring to light many of the key ethical responsibilities that engineers have in complex, hierarchical organizations. Second, his analysis is directly connected to one of the most notorious and influential automobile engineering ethics cases (a case that is also often discussed in the context of business ethics).

De George contends that engineers must uphold the (safety) standards of the time; such standards are a minimum threshold that their designs must not fall below (8). Furthermore, he believes that customers are entitled to know how much safety a car has. He offers a list of criteria for when it is: 1) morally permissible to report an issue to the public, and 2) when it would be morally obligatory. Yet in general he suggests that engineers should not be required to challenge managerial decisions (especially since doing so may put their career on the line). He argues that the primary responsibility for correcting such problems should fall on regulators and not engineers (a view we do not embrace).

Human beings, engineers included, have a natural (psychological) tendency to react to a "disaster" and then implement changes afterward. It can be challenging to garner the necessary will and resources towards solving a problem before the problem manifests itself. Furthermore, "failure" is often seen as what instructs designers in terms of creating a safer, "better" technology (24). At least some corporations adopt the mindset of waiting for regulators and/or legal liability to push them towards the implementation of a safer design (i.e., the same effect as De George's approach). However, as should be made abundantly clear by the deaths in the Pinto case, that type of attitude can cause significant harm to the public.

Technical and Other Related Complexities of Self-Driving Cars

The self-driving car can reveal several potential weaknesses inherent to De George's view and to traditional approaches to engineering challenges. The Pinto case and many other examples of automobile safety have focused on an individual component or on interrelated set of components often with a known fault (e.g., the GM ignition switch (25) or Takata airbags (26)). Much of the coverage of self-driving cars in the popular media has likewise been focused on components, particularly sensors for navigation and guidance, and on algorithms for safe driving (27).

A self-driving car, however, is an entire *system* at least part of which operates "autonomously." According to the 2016 SAE International standard J3016, the levels of automation of self-driving cars range from 0 to 5 as indicated in Table 1 (28).

Level 2 automation is already being incorporated into existing commercial vehicle brands including Mercedes, BMW, and Cadillac. The Tesla (Model S) incorporates Level 2 and some aspects of Level 3 automation which, as discussed later, has led to some accidents including at least one fatal incident. In this paper, we primarily seek to examine ethical issues related to Levels 3, 4, and 5. A key distinction among those levels is whether the "safety-critical driving functions" are fully entrusted to an automated system (Level 5) or whether a human being is supposed to retain control over those functions in at least some situations (Levels 3 and 4).

A myriad of technical complexities, some of which are described below, could interfere with the safety and reliability of self-driving cars. For example, a typical self-driving car is estimated to contain 100 million lines of code, which is approximately 10 times the amount of code in a fighter jet (13). Software testing has always been difficult (29); this kind of complexity makes it even more challenging. We should also keep in mind that it is not just the amount of code and its complexity that is

TABLE 1. Levels of Driving Automation (adapted from SAE [28]).

Human Driver Monitors Driving Environment	
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
Automated System Monitors Driving Environment	
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

worrisome; it is also the variety and uncertainty of situations that the system will face [30].

The complexity of a self-driving car's system architecture, including subsystems for the human interface, route planning, environment perception and modeling, and vehicle hardware actuators, all interconnected with a coordination and control module [31], could generate many outcomes that are difficult to anticipate. This is further complicated by the variability of the design pathways that different car manufacturers are pursuing. The volume of and interconnections between sensor data that have to be processed may (arguably) be a bigger problem that the sheer amount of code, especially given how such data must be processed in a short amount of time in order for a car to react promptly enough. Smooth and timely calibration across light detection and ranging (LIDAR) or other sensors is essential and difficult [32]. The associated complexity is increased by potential variables such as vehicle-to-vehicle (V2V) communication, cloud connectivity, and smart highways, all of which could intensify the need to process vast quantities of information almost instantaneously.

Users already seem to have a predisposition to develop an over-reliance on digital outputs.

Some of the advantages anticipated for automated cars are predicated on *all* vehicles being automated. If cars with human drivers are allowed to mingle with self-driving cars, an automated system will be much more difficult to design and test. However, even if human drivers are phased out, a collection of autonomous cars will still be part of a socio-technical system of enormous complexity. The design, development, and testing of swarms of robots is an area of research that is relatively young [33], but it seems clear that much progress will be required before a swarm of vehicles will be able to interact and operate safely [34]. The emergence of "normal accidents" (also known as "system accidents") is likely unavoidable due to the interactive complexity and tight coupling of the involved technical systems [35]. Intentional tampering (e.g., Volkswagen Diesel emissions tests) and hacking (e.g., Jeep Cherokee case) [36] are also legitimate sources of concern.

Another consideration is whether and how the Eliza Effect [37] might manifest itself; in other words, how might users deceive themselves in terms of a self-driving car's abilities? At times, an engineer's design choices directly contribute to the likelihood of a user's self-deception; for example, utilizing human-like features on a robot can lead users to inaccurately anthropomorphize the technology [38]. Users already seem to have a predisposition to develop an over-reliance on digital outputs as, for example, in the case of the Therac-25 radiation therapy machine [39]. Significant harm, and even death, can result from over-trust of computing technology, including in cases where it has led to airplane crashes [40]. Along these lines, a study by Robinette and colleagues indicates that participants may place too much trust in a robot during a simulated emergency situation even when the robot seems to be malfunctioning [41].

An added variable is that some companies are pursing design pathways that do not require the user to be actively involved in the operation of the car [42]. The underlying paternalistic logic of harm prevention may be well-intended, but numerous unintended, and potentially disastrous, consequences could result. While the list below highlights potential user-centered problems, they have a direct bearing on the designer's decisions and actions:

- Will the lack of control over the car cause a user to panic even when it is functioning normally?
- Which types of important information might the user miss? For example, will the user be attentive enough to notice if someone is trying to hack into the car?
- Will the user's driving skill diminish over time [43]?
- Will the user know how to respond if the car is "in trouble" especially if there is no steering wheel or other obvious means for intervening? Or if there is, would grabbing the wheel place the user (and others) at greater risk than letting the system handle the situation by itself?

With regard to the last point, a similar issue has emerged for airline pilots when they are relying on or interacting with an auto-pilot that may be malfunctioning [40], [44]. An overarching concern about normalization of deviance with regard to user behavior can certainly emerge as well [45]; in short, if they are not actively, cognitively engaged in the vehicle's operation, users will become less diligent about monitoring how it functions (to a point where it can easily be imagined that users could metaphorically if not literally become "asleep at the wheel"). These and numerous other considerations, integrally intertwined with the user's psychology and behavior, must be taken into account by designers.

Relying on Standards

Many scholars, including De George, would stipulate that the "standards of the time" can serve as a crucial means

for protecting the public from vehicle-related harms. While engineering standards are necessary, there are many occasions where they are not sufficient. This is saliently illustrated by the Pinto case where engineers satisfied rear-end collision and other safety standards of the time. Standards (whether from governmental entities and/or professional engineering societies) and regulations (federal and state) often have difficulty keeping pace with technological change, a challenge that is especially relevant to emerging technologies [46]. This seems to be occurring in the case of self-driving cars where regulations and standards have indeed been slow to materialize as the technology rapidly develops [47].

Arguably, standards for self-driving cars would have to be more rigorous than they are for traditional automobiles. Established "standards" for many of a car's safety features (e.g., front/rear impact tolerances) can be used to judge at least some of a designer's acts. But self-driving cars bring into the picture added variables for which standards must account; for example, a designer would have to determine how to prevent users from increasing the risk to which they expose themselves if the lack of control over the vehicle causes them to override automated systems.

There will be major challenges even in the most optimistic scenario, in which the cars will have standard interfaces that will encourage reliable interactions with each other and with a central system for coordination. In a less optimistic (but perhaps more realistic) scenario, automated cars will be developed by rival corporations that will be less interested in cooperation and more interested in keeping their competitive advantage. If designers are required to protect trade secrets and market advantage while developing, testing, and maintaining their separate automated car, then achieving system-wide reliability, and verifying that reliability, will be all the more difficult.

And as was previously mentioned, auto manufacturers are pursuing significantly different design pathways (e.g., Google vs. Mercedes). Thus, it will be difficult for regulators to develop a uniform approach to safety standards. Among the crucial divides in the self-driving car industry is whether a human being should remain in the driving loop at least to some degree (Levels 2-4) or whether the system should entirely take over the driving (Level 5). There are conflicting opinions in the engineering community about whether humans should be "artificially engaged" to keep their attention focused on a self-driving car's functioning or whether the car should be fully autonomous [48]. If the former is pursued, then considerations involving the interaction between the human operator and the autonomous system such as Mean Time Between Interventions (MTBI) and Mean Time to Intervene (MTTI) are crucial for designers to address [49].

Another significant divide is whether the safety and reliability of the car's functioning should be tied into an ongoing communication stream between the car and external systems (highway sensors, V2V, etc.) or whether the car should be "smart" enough so that it can operate independently from such input. This is sometimes referred to as the distinction between "connected" versus "automated" autonomous vehicles [50]. Coordination among vehicles that adhere to these distinct paradigms will be difficult.

An Appeal to Engineering Codes of Ethics?

One avenue for designers to obtain guidance on professional matters is through codes of ethics. The "paramountcy clause" from engineering codes (i.e., uphold "the safety, health, and welfare of the public") is certainly well-intentioned and important, but it can be unclear how to apply it to a particular case, especially when there is a professional difference of opinion or there is not much precedent on which to rely. Little specific guidance is provided thus far by professional codes regarding the design of self-driving cars or other "autonomous" technologies.

In general, beyond promulgating codes, professional societies might be reluctant to actively promote "aspirational" ethical behavior in part because they may lack consensus about which types of "good" behaviors should be openly endorsed. Conflicts between engineering priorities and business priorities may also limit the ability of professional societies to engage in ethics promotion and support [51]. Yet aspirational behavior is precisely what is needed in the case of self-driving cars given how much of an effect the cars will have on the lives and well-being of members of the public.

Moral Responsibility for Computing Artifacts: The Rules

"The Rules," championed by Keith Miller in collaboration with other computer scientists, engineers, and ethicists were created with the intent of providing guidance to the computing and engineering communities especially with respect to pervasive and autonomous technologies [9]. Unlike codes of ethics, The Rules do provide specific guidance relevant to the design of self-driving cars. The Rules are presented below with an accompanying commentary for each one regarding self-driving cars.

- Rule 1 — "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." (emphasis added)

This rule assigns moral responsibility to designers among others for "foreseeable effects." It is unclear, however, how predictable a self-driving car's (and its passenger's) behavior will be, especially in dynamic or unanticipated circumstances. What a designer can reasonably be expected to foresee is an ongoing source of debate, which is likely to be even more contentious with regards to emerging technologies such as self-driving cars. For example, "foreseeable use" and the designer's "intended use" are not necessarily the same thing [52]. Yet foreseeing how the user and other entities may interact with a self-driving car is particularly important, especially during testing phases; testing that ignores possible use cases will be far less effective.

- Rule 2 — "The shared responsibility of computing artifacts is not a zero-sum game. *The responsibility of an individual is not reduced simply because more people become involved* in designing, developing, deploying, or using the artifact. Instead, a person's responsibility includes being answerable for the behaviors of the artifact and for the artifact's effects after deployment, to the degree to which these effects are reasonably foreseeable by that person." (emphasis added)

The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, *and for the foreseeable effects of that artifact.*

Given that the creation of a self-driving car will result from the collective efforts of numerous individuals, many of the designers will largely be anonymous to users and the general public and perhaps even to their co-designers. Designers may be tempted to say their individual responsibility is "reduced" when the technology behaves in a less than optimal, and perhaps dangerous, manner because of how many people are involved in the design (often referred to as "the problem of many hands"), but that type of thinking might not be morally defensible.

Acknowledging that collective responsibility does not negate individual responsibility is critical. Designers and testers are part of a larger community of professionals who have ethical responsibilities for their decisions related to the self-driving car. Designers may experience much external pressure from manufacturers, or others, to weaken or ignore their responsibilities; yet they must seek to uphold the tenets of what it means to be an ethical professional, accepting their individual professional responsibilities.

- Rule 3 — "People who *knowingly use* a particular computing artifact are morally responsible for that use." (emphasis added)

This rule applies to users but the concept of "knowingly use" may be especially problematic in the case of self-driving cars; for example, when, where, and how to intervene may not be obvious to the human passengers of self-driving cars, especially in a crisis situation. Along related lines, how much knowledge about the technology's functioning is it reasonable to assume that users have? Moreover, how transparent will companies be about how the car is designed to behave when human users circumvent its safety features (e.g., a parent places a child in the car without supervision)?

- Rule 4 — "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*." (emphasis added)

Placing the self-driving car on the road is not merely a mundane, incremental step akin to introducing a newer model of automobile. Self-driving cars will be embedded in complex sociotechnical systems encompassing designers, manufacturers, drivers, motorcyclists, bicyclists, pedestrians, and regulators, as well as individual vehicles, roadways, and complex monitoring and control technologies. Interactions among drivers, passengers, pedestrians, vehicles, devices both internal to the car (such as GPS) and external (such as a sensor on the road or a building), and the external environment coalesce into the formation of a highly chaotic, difficult to predict system, especially considering how humans do not always act rationally and can have vastly different risk tolerances and behavioral patterns [53]–[54].

Moreover, not only will the technology of self-driving cars reshape the interaction between car and driver but its introduction will necessitate and be shaped by multifaceted social, legal, and political changes. Many macroethical factors will come into play (e.g., differing vehicle types, infrastructure planning, and environmental planning); many policy decisions will need to be made. For example, widespread use of self-driving cars could have a significant impact on urban planning due to the removal of parking spots [48]. In addition to vehicle safety requirements, regulations will be needed in terms of where the vehicles will be permitted to operate and whether a licensed driver must be in the car [55]. Along these lines, many individuals might not have a compelling need to obtain a driver's license and this can have far-reaching effects, including if one travels to a region that only has human-operated vehicles.

- Rule 5 — "People who design, develop, deploy, promote, or evaluate a computing artifact *should not explicitly or implicitly deceive users* about the artifact or its foreseeable effects, or about the sociotechnical systems in which the artifact is embedded." (emphasis added)

How transparent will companies be about how the self-driving car is designed to behave, especially in dynamic or dangerous traffic situations? Market forces and other forms of competition may pressure engineers and companies to present the car as "risk free" or "of minimal risk" to human drivers and passengers. This is already happening to some degree [56]. Yet Google has recently admitted that one of its cars could be blamed for an accident [57].

The first reported fatal accident involving a self-driving car occurred in May 2016. While under the control of its autopilot system, a Tesla car crashed into a tractor-trailer that was making a left turn in front of the car. A Tesla blog post suggested that since the tractor-trailer was white, it might not have been visible against the brightly lit sky to the car's autopilot system. That same post also stated, "Nonetheless, when used in conjunction with driver oversight, the data is unequivocal that Autopilot reduces driver workload and results in a statistically significant improvement in safety" [58].

Lucas Merian notes however that "The problem for Tesla has been that while its Autopilot … offered some (SAE standards) level 3 automation, there was no way to force a driver to retake control of the vehicle; that has resulted in several documented accidents" including a fatal one [59]. Merian goes on to argue that "(T)he problem … hasn't necessarily been that Tesla's Autopilot … isn't performing as promised, but that drivers place too much confidence in it and take their hands off the steering wheel and their attention from the road." Following the accident, Elon Musk asserted that Tesla is implementing changes to its autopilot system that will purportedly prevent this type of accident from recurring, including limits on how long a driver's hands can be off the wheel and improved radar for recognizing obstacles [60].

One could argue that Tesla's response to the accident is consistent with a number of "the Rules." The blogger's and Musk's statements seem to be offering a defense that Rule 5 is being upheld by the designers of the self-driving car as long as the public has a reasonably accurate view of the associated risks. The blogger also speaks to Rule 1 in so far as Tesla is claiming that it has considered the risks and is confident that their self-driving car reduces (although clearly does not eliminate) the chance of harm to the public. Rules 1 and 3 are both arguably addressed by Musk's remarks concerning recognition by Tesla that there is a driver in the loop who needs to be considered by the designers. Nevertheless, the fact that such accidents have occurred with automation in the range of Level 2–3 suggests, as we argue below, that vehicles in the range of Level 3–5 automation should not be permitted on the road until more thorough testing has been conducted.

Our Proposal

In the interest of public well-being and safety, designers, testers, managers, and others should sincerely engage with "The Rules" and contemplate their implications for decision making regarding self-driving cars. At an individual level, each designer should consider his/her ethical obligations in terms of creating safer technology. One approach that incorporates this type of thinking is value-sensitive design, which encourages designers to consider how the user's cherished values, such as autonomy, can be upheld while in the process of creating their technologies [61]–[62].

People who design, develop, deploy, promote, or evaluate a computing artifact *should not explicitly or implicitly deceive users* about the artifact or its foreseeable effects.

It is also essential that relevant professional communities become collectively involved in a deliberative and reflective process. More specifically, engineering, computing, and other communities should engage in a variety of activities related to anticipatory ethics, including how they can take steps to minimize the impact of system failures in self-driving cars. This could be akin to an "Asilomar-like" activity. Asilomar was a conference in 1975 where scientists gathered to discuss recombinant DNA research and then developed voluntary guidelines to help protect the public.

Unlike Asilomar, however, efforts concerning autonomous vehicles should be structured so as to include the views of stakeholders from outside of the science and engineering community [63]. This activity could be patterned after efforts being witnessed in other realms of emerging technology. For example, the BEINGS conference gathered together scientists, philosophers, lawyers, industry representatives, and others to discuss the ethics of gene editing technologies [64]. Stakeholders from across the globe have organized numerous events, including United Nations meetings, to address concerns about the use of military robots [65]–[66].

Following De George, customers are entitled to know about vehicle safety; this requires extraordinary transparency regarding self-driving cars due to technical complexities and inevitable tradeoffs occasioned by this new technology. Given that a broad and diverse base of users, with vastly different levels of education, may come to rely on the technology, legalistic and obtuse user agreements are unlikely to suffice.

How transparent will companies be about how the self-driving car is designed to behave, especially in dynamic or dangerous traffic situations?

In addition, we would argue that the makers of automated cars should be held accountable for their designs. Before any deployment, each company should demonstrate through carefully monitored trials on test driving tracks that the introduction of its system will not degrade road safety. This testing should take into account the issues raised above, including the interactions between competing brands of cars and with human drivers. While we anticipate that this requirement will add time to any eventual adoption of self-driving cars, we contend that this measure is appropriate considering the importance of protecting the public.

Requiring a demonstration of public safety before a product is released is not unprecedented. In fact, such demonstrations are routinely required of, for example, drug companies. We expect that automated cars may have more of an impact on public safety than any individual new drug or medical device. Therefore, detailed safety trials before deployment seems not only prudent but should be a minimum requirement. Unlike De George, we argue that responsibility for such safety trials should not rest primarily with managers and regulators; rather, for the reasons stated in this paper, we believe this shared responsibility should also be reflected in the ethical and professional behavior of engineering designers of self-driving cars, even in the face of external pressure from management or other entities. The fact that such requirements have not heretofore been enforced, since cars with significant levels of automation are already on public roads, suggests that in the case of automated cars, economic forces and technological momentum have superseded the public good; this inversion of values should be halted and reversed.

References

[1] Y. Berra and D. Kaplan, *When You Come to a Fork in the Road, Take It!: Inspiration and Wisdom From One of Baseball's Greatest Heroes*. New York, NY: Hyperion, 2002.

[2] E. Behrmann, "VW stops the music as diesel scandal buries culture of spending," *Automotive News*, Oct, 6, 2015; http://www.autonews.com/article/20151006154644/OEM02/151009883, accessed June 15, 2016.

[3] D. Gelles, H. Tabuchi, and M. Dolan, "Complex car software becomes the weak spot under the hood," *New York Times*, Sept. 26, 2015.

[4] N.A. Greenblatt, "Self-driving cars will be ready before our laws are," *IEEE Spectrum*, Jan. 16, 2016, http://spectrum.ieee.org/transportation/advanced-cars/selfdriving-cars-will-be-ready-before-our-laws-are, accessed Mar. 1, 2016.

[5] P. Lin, "The ethics of autonomous cars," *The Atlantic*, Oct. 8, 2013, http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/, accessed Apr. 18, 2016.

[6] M.N. Mladenovic and T. McPherson, "Engineering social justice into traffic control for self-driving vehicles," *Science and Engineering Ethics*, Aug. 1, 2015 (Epub version).

[7] J.H. Moor, "Why we need better ethics for emerging technologies," *Ethics and Information Technology*, vol. 7, pp. 111–119, 2005.

[8] R.T. De George, "Ethical responsibilities of engineers in large organizations: The Pinto case," *Business & Professional Ethics Journal*, pp. 1-14, 1981.

[9] K. Miller, "Moral responsibility for computing artifacts: 'The Rules'," *IT Professional*, pp. 57-59, May/June 2011; http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5779006, accessed Feb. 4, 2016.

[10] P.A.E. Brey, "Anticipatory ethics for emerging technologies," *NanoEthics*, vol. 6, no. 1, pp. 1-13, 2012.

[11] H. Sutcliffe, "A report on responsible research & innovation," prepared for DG Research and Innovation, European Commission, 2011; https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/dae-library/a_report_on_responsible_research_innovation.pdf, accessed Apr. 19, 2016.

[12] M. Montgomery, "The new big 4 of the auto world: Tesla, Google, Apple and Uber," *Forbes.com*, Nov. 18, 2015; http://www.forbes.com/sites/mikemontgomery/2015/11/18/meet-the-new-big-4-of-the-auto-world-tesla-google-apple-and-uber/, accessed June 3, 2016.

[13] E. Taylor, "Testing of software adds to urgency in race for driverless cars," *Reuters*, Mar. 27, 2015; http://www.reuters.com/article/us-autos-driverless-idUSKBN0MN1E820150327, accessed Feb. 4, 2016.

[14] National Highway Traffic Safety Administration (NHTSA), *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, 2015; http://www-nrd.nhtsa.dot.gov/pubs/812115.pdf, accessed Apr. 20, 2016.

[15] Association for Safe International Road Travel, *Annual Global Road Crash Statistics;* http://asirt.org/initiatives/informing-road-users/road-safety-facts/road-crash-statistics, accessed Apr. 20, 2016.

[16] D. Freeman, "Self-Driving cars could save millions of lives – But there's a catch," *HuffPost Tech*, Feb. 18, 2016; http://www.huffingtonpost.com/entry/the-moral-imperative-thats-driving-the-robot-revolution_us_56c22168e4b0c3c550521f64, accessed Mar. 1, 2016.

[17] H. King, "Google: Human drivers are the problem," *CNN.com*. May 12, 2015; http://money.cnn.com/2015/05/12/autos/google-self-driving-cars-accidents/, accessed Apr. 18, 2016.

[18] Eno Center for Transportation, *Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations*, 2013, https://www.enotrans.org/wp-content/uploads/2015/09/AV-paper.pdf, accessed Feb. 4, 2016.

[19] D. Howard, "Robots on the road: The moral imperative of the driverless car," *Science Matters*, 2013; http://donhoward-blog.nd.edu/2013/11/07/robots-on-the-road-the-moral-imperative-of-the-driverless-car/#.U1oq-1ffKZ1, accessed Apr. 20, 2016.

[20] B. Vlasic, "U.S. proposes spending $4 billion on self-driving cars," *New York Times*, Jan. 14, 2016; http://www.nytimes.com/2016/01/15/business/us-proposes-spending-4-billion-on-self-driving-cars.html, accessed June 3, 2016.

[21] A. Hevelke and J. Nida-Rümelin, "Responsibility for crashes of autonomous vehicles: An ethical analysis," *Science and Engineering Ethics*, vol. 21, pp. 619-630, 2015.
[22] L.D. Riek and D. Howard, "A code of ethics for the human-robot interaction profession," in *Proc.We Robot*, 2014.
[23] N.J. Goodall, "Can you program ethics into a self-driving car?," *IEEE Spectrum*, June 2016; http://spectrum.ieee.org/transportation/self-driving/can-you-program-ethics-into-a-selfdriving-car, accessed June 3, 2016.
[24] H. Petroski, *To Engineer Is Human: The Role of Failure in Successful Design*. Vintage, 1992.
[25] T. Basu, "Timeline: A history of GM's ignition switch defect," *NPR*, Mar. 31, 2014; http://www.npr.org/2014/03/31/297158876/timeline-a-history-of-gms-ignition-switch-defect, accessed June 7, 2016.
[26] National Highway Traffic Safety Administration (NHTSA), *Recalls Spotlight: Takata Air Bags Recalls;* http://icsw.nhtsa.gov/safercar/rs/takata/, accessed June 7, 2016.
[27] B. Schweber, "The autonomous car: A diverse array of sensors drives navigation, driving, and performance," Mouser Electronics, n.d.; http://www.mouser.com/applications/autonomous-car-sensors-drive-performance, accessed June 3, 2016.
[28] SAE International, "Automated driving: levels of driving automation are defined in new SAE International Standard J3016;" http://www.sae.org/misc/pdfs/automated_driving.pdf, accessed Dec. 6, 2016.
[29] J.A. Whittaker, "What is software testing? And why is it so hard?," *IEEE Software,* vol. 17, no. 1, pp. 70-79, 2000.
[30] M. Althoff and A. Mergel, "Comparison of Markov chain abstraction and Monte Carlo simulation for the safety assessment of autonomous cars," *IEEE Trans. on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1237-1247, 2011.
[31] V. Gummadi, "An algorithm for driverless cars in India," *funmonk,* Feb. 26, 2014; http://www.funmonk.co/2014/02/26/an-algorithm-for-driverless-cars-in-india/, accessed Feb. 4, 2016.
[32] R.W. Wolcott and R.M. Eustice, "Visual localization within LIDAR maps for automated urban driving," in *Proc. 2014 IEEE/RSJ Int. Conf. Intelligent Robots and Systems* (IROS 2014), 2014, pp. 176-183.
[33] A.F.T. Winfield, C.J. Harper, and J. Nembrini, "Towards dependable swarms and a new discipline of swarm engineering," *Swarm Robotics*. Germany: Springer Berlin Heidelberg, 2004, pp. 126-142.
[34] J.P. Hecker and M.E. Moses, "Beyond pheromones: Evolving error-tolerant, flexible, and scalable ant-inspired robot swarms," *Swarm Intelligence,* vol. 9, no. 1, pp. 43-70, 2015.
[35] C. Perrow, *Normal Accidents: Living with High-Risk Technologies*, 2nd ed. Princeton, NJ: Princeton Univ. Press, 1999.
[36] A. Greenberg, "Hackers remotely kill a jeep on the highway—With me in it," *Wired*, July 21, 2015; http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/, accessed Feb. 4, 2016.
[37] S. Turkle, *Life on the Screen: Identity in the Age of the Internet*. New York, NY: Simon and Schuster, 1995.
[38] M.A. Boden, "Robots and anthropomorphism," in Tech. Rep. WS-06-09. Menlo Park, CA: AAAI, 2006, pp. 69-74.
[39] B.M. O'Connell and J.R. Herkert, "Engineering ethics and computer ethics: Twins separated at birth?," *Techné,* vol. 8, no.1, Fall 2004; http://scholar.lib.vt.edu/ejournals/SPT/v8n1/oconnell.html, accessed Feb. 4, 2016.
[40] N. Carr, *The Glass Cage: Automation and Us*. Norton, 2014.
[41] P.R. Robinette *et al.*, "Overtrust of robots in emergency evacuation scenarios," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction (HRI 2016)*, Christchurch, New Zealand, 2016, pp. 101-108.
[42] J. Markoff, "Google's next phase in driverless cars: No steering wheel or brake pedals," *New York Times,* May 27, 2014; http://www.nytimes.com/2014/05/28/technology/googles-next-phase-in-driverless-cars-no-brakes-or-steering-wheel.html, accessed June 14, 2015.
[43] S.M. Casner, E.L. Hutchins, and D. Norman, "The challenges of partially automated driving," *Commun. ACM*, vol. 59, no. 5, pp. 70-77, 2016.
[44] D.A. Mindel, *Our Robots, Ourselves: Robotics and the Myths of Autonomy*. Viking, 2015.
[45] D. Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago, IL: Univ. of Chicago Press, 1997.
[46] G.E. Marchant, B.R. Allenby, and J.R. Herkert, Eds. *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*. Springer, 2011.
[47] L. Tillemann and C. McCormick, "This could be the biggest hurdle for driverless cars," *Fortune.com*, Feb. 15, 2016; http://fortune.com/2016/02/15/driverless-cars-google-lyft/, accessed June 3, 2016.
[48] R. Eustice, *University of Michigan's Work Toward Autonomous Cars*, 2015; http://www.umtri.umich.edu/sites/default/files/Ryan.Eustice.UM_.Engineering.IT_.2015B.pdf, accessed Apr. 20, 2016.
[49] T.W. Fong *et al.*, "A preliminary study of peer-to-peer human-robot interaction," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics (SMC '06),* 2006, vol. 4, pp. 3198-3203.
[50] European Commission, *GEAR 2030 Discussion Paper: Roadmap on Highly Automated Vehicles*, 2016; https://circabc.europa.eu/sd/a/a68ddba0-996e-4795-b207-8da58b4ca83e/Discussion%20Paper%20-%20Roadmap%20on%20Highly%20Automated%20Vehicles%2008-01-2016.pdf, accessed June 8, 2016.
[51] J.R. Herkert, "Future directions in engineering ethics research: Microethics, macroethics and the role of professional societies," *Science and engineering ethics,* vol. 7, no. 3, pp. 403-414, 2001.
[52] J.R. Herkert, "Professional societies, microethics, and macroethics: Product liability as an ethical issue in engineering design," *Int. J. Engineering Education,* vol. 19, no. 1, pp. 163-167, 2003.
[53] J. Brännmark and N-E Sahlin, "Ethical theory and the philosophy of risk: First thoughts," *J. Risk Research,* vol. 13, no. 2, pp. 149-161, 2010.
[54] P. Slovic and E. Peters, "Risk perception and affect," *Current Directions Psych. Sci.,* vol. 15, no. 6, pp. 322-325, 2006.
[55] D. Shepardson and P. Lienert, "Exclusive: In boost to self-driving cars, U.S. tells Google computers can qualify as drivers," *Reuters*, Feb. 10, 2016; http://www.reuters.com/article/us-alphabet-autos-selfdriving-exclusive-idUSKCN0VJ00H, accessed Apr. 20, 2016.
[56] M. Fox, "Self-driving cars safer than those driven by humans: Bob Lutz," *CNBC*, Sept. 8, 2014; http://www.cnbc.com/2014/09/08/self-driving-cars-safer-than-those-driven-by-humans-bob-lutz.html, accessed June 9, 2016.
[57] A. Davies, "Google's self-driving car caused its first crash," *Wired*, Feb. 29, 2016; http://www.wired.com/2016/02/googles-self-driving-car-may-caused-first-crash/, accessed Apr. 19, 2016.
[58] M. McFarland, "Tesla's autopilot probed by government after crash kills driver," *CNN Money,* July 1, 2016; http://money.cnn.com/2016/06/30/technology/tesla-autopilot-death/index.html, accessed Sept. 12, 2016.
[59] L. Merian, "Ford remains wary of Tesla-like autonomous driving features," *Computerworld,* Aug. 19, 2016; http://www.computerworld.com/article/3109217/car-tech/ford-wary-of-tesla-like-autonomous-driving-features.html, accessed Dec. 12, 2016.
[60] N.E. Boudette, "Elon Musk says pending Tesla updates could have prevent fatal crash," *New York Times*, Sept. 11, 2016.
[61] B. Friedman, "Value-sensitive design," *Interactions,* vol. 3, no. 6, pp. 16-23, 1996.
[62] B. Friedman, P.H. Kahn, Jr., and A. Borning, "Value sensitive design and information systems," in *Human-Computer Interaction and Management Information Systems,* P. Zhang and D. Galletta, Eds.New York, NY: Sharpe, 2006, pp. 348-372.
[63] J.B. Hurlbut, "Limits of responsibility: Genome editing, Asilomar, and the politics of deliberation," *Hastings Center Rep.,* vol. 45, no. 5, pp. 11-14, 2015; http://onlinelibrary.wiley.com/doi/10.1002/hast.484/pdf, accessed June 3, 2016.
[64] Biotechnology and the Ethical Imagination: A Global Summit (BEINGS), 2015; http://www.beings2015.org, accessed June 7, 2016.
[65] K.D. Atherton, "The international community is about to debate killer robots," *Popular Sci.*, Apr. 11, 2016; http://www.popsci.com/international-community-is-about-to-debate-killer-robots, accessed June 9, 2016.
[66] "UN meeting targets 'killer robots'," *UN News Centre*, May 14, 2014; http://www.un.org/apps/news/story.asp?NewsID=47794, accessed June 9, 2016.

TS