Introduction

The inevitable occurred on March 18, 2018, when an autonomous vehicle owned and operated by Uber Technologies, Inc. struck and killed Elaine Herzberg as she crossed a street with her bicycle. Her friends describe her as just emerging from homelessness. A human driver was in the sports utility vehicle, but the SUV was in autonomous mode when it struck Herzberg.[1] Video footage indicates it was night and Herzberg was not in a crosswalk. The footage also suggests the car did not apply brakes or otherwise attempt to avoid hitting Herzberg. A preliminary report of the National Transportation Safety Board revealed that the vehicle's self-driving system first detected Herzberg about six seconds before impact. As the car and Herzberg converged, the autonomous system classified Herzberg as an unknown object, next as a vehicle and then as a bicycle. Just 1.3 seconds before impact, the system concluded that an emergency braking maneuver was required; however, such maneuvers were not enabled while the vehicle was under computer control. This was to prevent "erratic vehicle behavior." The system was not designed to alert the human operator.[2]

Others had died in accidents involving self-driving cars before this incident,[3] but this was the first time someone with no voluntary connection to an autonomous vehicle had been killed by one. Immediately, discussions arose in the comments sections of news reports and via Twitter posts about who should be held responsible for Herzberg's death. Was it the car? Uber? Herzberg herself? The human driver who was there to override the SUV's systems in an emergency? (An accident report by the Tempe Police Department indicated that

[1] Daisuke Wakabayashi, *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam*, N.Y. Times, Mar. 19, 2018, at A1.

[2] U.S. Nat'l Transp. Safety Bd., Preliminary Report: Highway, HWY18MH010 (2018), at 2 [hereinafter NTSB Preliminary Report].

[3] A recent article from the popular press is Adrienne LaFrance, *Can Google's Driverless Car Project Survive a Fatal Accident?* Atlantic, Mar. 1, 2016, http://www.theatlantic.com/technology/archive/2016/03/google-self-driving-car-crash/471678/. The first death in a self-driving vehicle occurred on May 7, 2016. See Bill Vlasic & Neal E. Boudette, *As U.S. Investigates Fatal Tesla Crash, Company Defends Autopilot System*, N.Y. Times, Jun. 16, 2016, http://www.nytimes.com/2016/07/13/business/tesla-autopilot-fatal-crash-investigation.html?_r=0.

the driver had been streaming a video program while in the car.[4]) The state of Arizona for failing to adequately regulate autonomous vehicles? (I will return to these facts and questions in Chapter 2.)

As that event was making headlines, another, perhaps more far-reaching development was coming to light. This was the news that information from millions of Facebook users had been acquired, without permission, by Cambridge Analytica, which had in turn used that information to construct behavioral models as it advised clients who wanted to frame and direct public opinion.[5] It was possible that as many as 87 million users had been affected.[6] There was debate about whether such efforts to sway the public succeeded, but at least some claimed that they had influenced important public decisions, including Brexit and the 2016 US election. Again, questions were raised about who should be blamed for the invasion of privacy, and again, there were likely culprits: Cambridge Analytica and its principals, of course, but also Facebook for collecting private data in the first place and then failing to protect it.

If nothing else, the loss of Herzberg's life and the Cambridge Analytica scandal announced that autonomous machines and systems and the harms they can cause are no longer a distant prospect. The issue of how law can be used to prevent them from harming others and how to assign responsibility if they do is fast becoming ripe. Policymakers, industry leaders, and legal scholars have seen this day coming for some time and there is a growing literature on the role law can and should play as artificial intelligence and the technologies it powers become a greater part of everyday experience.

This book contributes to that discussion by plotting a possible trajectory for the relationship between the law and autonomous machines and systems. That trajectory begins where it often does when something new appears on the scene: communities use already existing legal doctrines and principles to comprehend and respond to it. This in turn leads to debate about how well suited these doctrines and principles are to new developments and whether there will be a need for significant changes in them.

---

[4] Associated Press, *Police: Backup Driver in Fatal Uber Crash was Distracted*, N.Y. Times, June 22, 2018, https://www.nytimes.com/aponline/2018/06/22/us/ap-us-uber-autonomous-vehicle.html.

[5] Matthew Rosenberg, Nicholas Confessore & Carole Cadwallader, *How Trump Consultants Exploited the Facebook Data of Millions*, N.Y. Times, Mar. 17, 2018, https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.

[6] Cecilia Kang & Sheera Frenkel, *Facebook Now Says Improper Data Use Affected 87 Million*, N.Y. Times, Apr. 4, 2018, https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&module=first-column-region&region=top-news&WT.nav=top-news.

My thesis is that the growing sophistication of machines and systems creates impulses for legal norms and machines to coevolve. In a sense, this coevolution is already happening in scholarship and within the industry along two parameters and lines of development. One is the nexus between the machine and human activity. At present, the legal system is trying as much as possible to associate the actions of artificial agents and their consequences to individuals or groups of human beings, and at this point the technology is still limited enough for this approach to remain viable. At this point it is human beings who use technology as tools, and it is those human beings who are ultimately responsible for any harms caused by those tools. The doctrines being used include individual tort liability, product liability, agency, joint criminal enterprise, aiding and abetting, conspiracy, and command responsibility. With modifications, such doctrines seem to work relatively well for less sophisticated machines and more or less so in cases where sophisticated machines are clearly carrying out the will of humans.

However, this is where the second parameter, the degree of autonomy of the machine or system as decision maker, makes itself felt. Although some commentators argue that autonomous technologies will always remain a tool used by human beings, others believe that the more autonomy machines and systems achieve, the more tenuous becomes the strategy of attributing and distributing legal responsibility for their behavior to human beings. Strict liability is of course always available, but the law tends to be more comfortable with assigning legal liability to someone when he is personally culpable for a harm and far less so with liability or guilt by association. In this sense, the law corresponds to prevailing views of individual moral responsibility.

Thus, as machines and systems reach higher levels of autonomy, the effectiveness of specific proposals such as "use product liability and other tort doctrines if a patient is harmed by the use of nanotechnology" or "apply the doctrine of command responsibility if an autonomous weapon 'commits' an act that would constitute a war crime if a human committed it," will depend in part on our comfort with solutions to the problem of associational responsibility. Even in cases where autonomous technologies are clearly being used or supervised by human beings, some of the incentives created by legal rules, such as the incentive to take due care, weaken because humans will have less control over truly independent technologies. Further, since a machine or system cannot at this point feel legal sanctions, other purposes of the law, particularly those that motivate criminal law, are thwarted. As a result, we may be forced to become more comfortable with group legal responsibility or responsibility by association, or face the prospect of manufacturers, owners, or users of such machines and systems becoming insulated from legal responsibility.

For these reasons, one would expect two things to occur. First, a growing awareness of how permeable the concept of responsibility is, in part because

technology itself has the potential to affect the way we understand ourselves and our own agency, could make legal doctrines based on associational responsibility more acceptable to us, and alternative forms of redress or compensation for harm, such as insurance, might gain greater importance. I argue, however, that although we might be open to some changes to the rules of responsibility, the idea of personal responsibility and the legal doctrines informed by it will likely persist. Second, our understandings of responsibility will likely join with other incentives to design even more sophisticated artificial agents. We would expect to see designers try to instill a sense of legal responsibility within the machine itself. Of course, as just discussed, machines and systems are not cognizant of the law, far less do they subjectively appreciate or value it. For now, all we can do is program the machines to act as much as possible in conformity to the law, for example, by instructing autonomous cars to obey traffic laws or autonomous weapons to obey the law of war. But this raises a question whether law can always be reduced to rules of decision in the settings in which we expect autonomous machines and systems to operate. In addition, many of the legal issues involving autonomous machines will be retrospective in nature: we will need to determine *ex post* whether an artificial agent's action has legal significance. As we will see, however, these questions are not keeping designers from trying to create autonomous technologies that conform to law.

The need to instill machines and systems with a sense of law will vary according to their level of sophistication, but over the long term, technologies at the highest levels of autonomy will need to be programmed so they are "motivated" to engage in the kinds of pro-social behaviors the law is designed to promote. The case of HAL in *2001: A Space Odyssey* and critiques of Asimov's laws of robotics suggest this can succeed only to some extent, but as autonomous technologies gain those kinds of prosocial capacities, this will strengthen calls already being sounded to grant autonomous machines and systems legal and moral rights. Some are already urging that autonomous technologies be given legal personhood to satisfy third parties who have been harmed by them while at the same time avoiding some of the problems raised by associative responsibility. If this happens, the world will of course be a very different place, and the law will have gained another subject.

This book is divided into three parts. Part I, comprising Chapters 1 and 2, surveys briefly the issue of autonomous machines and the existing legal approaches that frame and address the problem. Developments are occurring at almost bewildering speed, so Chapter 1 identifies only basic trends. Those trends point to a future in which artificial intelligence will be ubiquitous, not only with respect to our forms of transportation, but more importantly behind the scenes as significant aspects of everyday life are impacted by artificial intelligence. In this context, law will need to address large and complex

systems of humans and machines who work together. Chapter 2 discusses how commentators are using existing legal concepts taken from torts, contracts, and international law to respond to the issue of harm and reviews the debate about the extent to which such doctrines in their current forms can address the situations that will arise when autonomous machines become more prevalent.

Part II turns to the relationship between law and ethics. To the extent that the legal doctrines discussed in Chapter 2 do not adequately address harms caused by autonomous machines, it is in large part because of the law's discomfort with associative responsibility, a discomfort shared and informed by most of the literature on moral responsibility. Chapter 3 observes that the legal approaches currently used to frame and address harms caused by autonomous machines focus primarily on individual legal responsibility. Even applications of the law to groups still tend to frame their analyses in individualistic terms. This dovetails with generally accepted understandings of moral responsibility, which can be traced from Aristotle to the present day. This raises the question whether approaches designed with the individual in mind are well suited to address large systems which will produce and employ autonomous machines.

Chapter 4 turns to the problem of group responsibility. Within ethics, the literature on the moral responsibility of groups is most relevant to the problems of associative responsibility. That literature provides some guidance on whether it is coherent to ascribe responsibility to groups; if so, which types of groups might be subject to responsibility; how the responsibility of the group can be distributed to its members; and the "pragmatics" of ascribing responsibility to groups. At the same time, in part because of differences between law and ethics, and because of the nature of the problem, the literature does not provide completely satisfactory answers. Concepts from complexity theory suggest that the problem might be intractable. This is because the "behaviors" of groups might be the nonlinear, emergent phenomena that arise from the complex interactions of individuals and subgroups. This creates an argument that it is hard, if not impossible, to trace causal lines between the actions of individuals and what happens at the group level. In such situations, any responsibility we attribute to an individual for what happens at the group level would necessarily be a fiction. The chapter will conclude by discussing the relevance of group responsibility and complexity theory to the problem of autonomous machines and systems.

Part III explores further how one could reexamine our current understandings of responsibility in light of autonomous machines and systems. The concern there might be that gaps between harms caused by autonomous technologies and existing legal and moral concepts of responsibility lead to two interweaving "strategies" to narrow those gaps. Chapter 5 discusses the first: to refine or alter the concept of responsibility. Such attempts come from a number of perspectives. One is to rely more on concepts that underlie strict

liability. Another is to focus on the distinction made by some contemporary ethical theorists between responsibility as identifying an agent as contributing to a particular harm on the one hand and having the agent suffer consequences because of that contribution on the other. This is amenable to another approach that centers more on the victim of harm than on the perpetrator. An approach that centers more on compensating the victim provides a natural segue to compensation systems like commercial and social insurance. It is almost certain that insurance will be used to compensate injured parties and will play an important role in responding to harm caused by autonomous machines. At the same time, insurance has its limitations because of the problem of moral hazard, it does not perform the more punitive or retributive functions of holding someone responsible for a harm, and there are some limits to insurance as a means of pooling risk. The chapter will discuss these approaches and assess the strengths and weaknesses of insurance schemes as they apply to autonomous machines.

Chapter 6 turns to the second strategy to narrow gaps in responsibility, this time from the human perspective. One way to alter concepts of responsibility is to extend the boundaries of agency to include humans and artificial agents together, thus raising again the issues posed by group responsibility. However, those modifications require the expansion of ethical and legal subjects capable of bearing responsibility that many individual human beings will find hard to accept, although some change might be possible at the margins. The chapter will discuss those approaches and support my argument that they are ultimately unworkable on a large scale.

Part IV moves from a focus on the human being as the locus of responsibility to autonomous machines and systems. The impasses discussed at the conclusions of Chapters 5 and 6 (as well as the simple desire to avoid liability) motivate in part the second strategy to close possible gaps between machines and harms caused by them: to reduce harm by designing autonomous machines and systems that "obey" the law. This is the topic of Chapter 7. As discussed above, at this point of course, autonomous technologies are not cognizant of the law, far less do they appreciate or value it in the subjective sense; all we can do is program machines and systems to operate in ways that conform to the law. The chapter will discuss ways in which machines and systems are being designed to comply with the law and assess the debate about the extent to which this strategy can be successful. It will also point out that if this effort succeeds, it will not be surprising if autonomous technologies, not human beings, set appropriate standards of care.

Chapter 8 projects further into the future. It discusses ways in which autonomous technologies might be designed to exhibit prosocial behaviors and to have systems of ethics. This leads to Chapter 9, which will discuss efforts to

give autonomous machines and systems legal personality and explore whether they should be given legal rights and moral consideration.

Part IV puts the discussion into perspective. Chapter 10 concludes the book by arguing that to some extent, the trajectory of the coevolution of legal responsibility and autonomous machines laid out in Chapters 5 through 9 needs to be cabined. As will be seen in Chapters 2 and 3, the law already applies to complex systems, albeit with concepts borrowed heavily from individual legal and moral responsibility. It is only if society feels it is necessary to become finer grained in assigning responsibility, to move from largescale entities that design and manufacture autonomous machines and systems to individual designers and engineers who could be said to have contributed to the defects that led to harms and to individuals in the chain of command who use autonomous machines, that the problems of associational responsibility discussed in this article become more salient. Autonomous technologies would then set us along that trajectory, if by that time their decision-making capacity is so sophisticated it is hard to attribute responsibility for harms they cause to their human coworkers, supervisors, or those who designed them, but they are not autonomous enough to merit legal, let alone moral agency, so that they can be meaningfully blamed and punished for what they have done. It ends by identifying points along the trajectory that will merit careful attention by jurists and policymakers.