2016

Technological Opacity, Predictability, and Self-Driving Cars

Harry Surden
University of Colorado Law School

Mary-Anne Williams
University of Technology Sydney

TECHNOLOGICAL OPACITY, PREDICTABILITY, AND SELF-DRIVING CARS

Harry Surden† *& Mary-Anne Williams*†

Autonomous or "self-driving" cars are vehicles that drive themselves without human supervision or input. Because of safety benefits that they are expected to bring, autonomous vehicles are likely to become more common. Notably, for the first time, people will share a physical environment with computer-controlled machines that can both direct their own activities and that have considerable range of movement. This represents a distinct change from our current context. Today people share physical spaces either with machines that have free range of movement, but are controlled by people (e.g. automobiles) or with machines that are controlled by computers, but highly constrained in their range of movement (e.g. elevators). The movements of today's machines are thus broadly predictable. The unrestricted, computer-directed movement of autonomous vehicles is an entirely novel phenomenon that may challenge certain unarticulated assumptions in our existing legal structure.

Problematically, the movements of autonomous vehicles may be less predictable to the ordinary people who will share their physical environment—such as pedestrians—than the comparable movements of human-driven vehicles. Today, a great deal of physical harm that might otherwise occur is likely avoided through humanity's collective ability to predict the movements of other people. In anticipating the behavior of others, we employ what psychologists call a "theory of mind." Theory of mind cognitive mechanisms allow us to extrapolate from our own internal mental states in order to estimate what others are thinking or likely to do. These cognitive systems allow us to make instantaneous, unconscious judgments about the likely actions of people around us, and therefore, to keep ourselves safe in the driving context. However, the theory of mind mechanisms that allow us to accurately model the minds of other people and interpret their communicative signals of attention and

† Professor of Law, University of Colorado Law School

† Professor of Engineering and Robotics, University of Technology Sydney. We would like to thank Andrew Coan, Seema Shah, Carol Rose, Jane Bambauer, Larry Head, Eric Frew, David Levine, Bernard Chao, Blake Reid, Viva Moffat, Kristelia Garcia, Susan Nevelow-Mart, Colter Donahue, Helen Norton, Ashkan Soltani, Margot Kaminski, Sam Arbseman, and Ryan Calo for their helpful comments and suggestions. We would also like to thank the University of Colorado Law School, Stanford Law School CodeX Center, and the University of Technology Sydney for their support.

intention will be challenged in the context of non-human, autonomous moving entities such as self-driving cars.

This Article explains in detail how self-driving vehicles work and how their movements may be hard to predict. It then explores the role that law might play in fostering more predictable autonomous moving systems such as self-driving cars, robots, and drones.

TABLE OF CONTENTS

INTRODUCTION

As we navigate an environment filled with moving people and automobiles, how do we avoid injuring one another? The ability to predict the actions of others plays a crucial role.[1] Consider a pedestrian about to enter a crosswalk in front of an approaching car.[2] Before stepping in front of the car, the pedestrian must predict whether the driver is likely to stop. As part of this decision, the pedestrian will make a series of instantaneous observations about the driver's perceptions, capabilities, and intentions: Does the driver see the pedestrian? Is the driver capable of braking? Is the driver planning to stop?

In addition to observation, we often rely upon communication to predict the behavior of others.[3] Pedestrians and drivers sometimes make eye contact, silently indicating awareness of each other's presence. In

1 *See* MARC GREEN ET AL., FORENSIC VISION WITH APPLICATION TO HIGHWAY SAFETY 335–36 (3d ed. 2008).

2 For the sake of this example, assume that the crosswalk has no traffic signal or stop sign, so that the car will not necessarily need to stop in the absence of a pedestrian.

3 *See* Adam Kendon, *Introduction: Current Issues in the Study of "Nonverbal Communication"*, *in* NONVERBAL COMMUNICATION, INTERACTION, AND GESTURE: SELECTIONS FROM SEMIOTICA 1, 18 (Adam Kendon, Thomas A. Sebeok & Jean Umiker-Sebeok eds., 1981) (stating that in a crowd, pedestrians avoid collisions by looking at one another to detect "information about each other's direction of movement from the movement of their bod[ies]"). For an example of non-verbal driver communication, see Burkhard Bilger, *Auto Correct: Has the Self-Driving Car at Last Arrived?*, NEW YORKER (Nov. 25, 2013), http://www.newyorker.com/magazine/2013/11/25/auto-correct ("[N]udging [a car forward at an intersection] is a kind of communication. . . . It tells people that it's your turn.").

other scenarios, a driver might explicitly wave a pedestrian to cross, or visibly reduce speed, in order to communicate his intention to stop.[4] With each communication, the parties gain more information and are able to more reliably assess one another's intentions.[5] A pedestrian who receives a wave to cross from a driver in a visibly slowing car, can enter the crosswalk confident that the driver will stop.

Our ability to predict the actions of others is a more general phenomenon beyond the automobile setting. A great deal of physical harm that might otherwise occur is avoided through humanity's collective ability to anticipate the movements of others and react accordingly.[6] In anticipating the behavior of other people we employ what psychologists call a "theory of mind."[7] The term "theory of mind" refers to our ability to extrapolate from our own internal mental states to estimate what others are thinking, feeling, or likely to do.[8]

Theory of mind cognitive mechanisms allow us to make instantaneous, unconscious judgments about the likely actions of those around us in order to keep ourselves safe.[9] Imagine the earlier pedestrian observing a driver who is looking down at his cell phone. The pedestrian will intuitively understand that the distracted driver has probably not seen her and can avoid stepping into the crosswalk. By putting ourselves in the position of others, and assessing what they do (or do not) know, we can often anticipate their actions and preserve our own safety.[10] More broadly, although law creates incentives to reduce harm, society also implicitly relies on such cognitive-social mechanisms to avoid injuries that might otherwise occur as people and vehicles move about in the same physical space.[11]

---

4 GREEN ET AL., *supra* note 1, at 335.

5 Brenda Ocampo & Ada Kritikos, *Interpreting Actions: The Goal Behind Mirror Neuron Function*, 67 BRAIN RES. REVIEWS 260 (2011) (discussing using observation of others to determine intentions).

6 *See* GREEN ET AL., *supra* note 1, at 336 ("Drivers reacted to several pedestrian cues, including location in the road, proximity to the road . . . and even slowing.").

7 Jonathan Vitale, Mary-Anne Williams & Benjamin Johnston, *Socially Impaired Robots: Human Social Disorders and Robots' Socio-Emotional Intelligence*, *in* SOCIAL ROBOTICS: 6TH INTERNATIONAL CONFERENCE, ICSR 2014, SYDNEY, NSW, AUSTRALIA, OCTOBER 27–29, 2014 PROCEEDINGS 350, 351 (Michael Beetz, Benjamin Johnston & Mary-Anne Williams eds., 2014); Chris D. Frith & Uta Frith, *How We Predict What Other People Are Going to Do*, 1079 BRAIN RES. 36, 41 (2006).

8 For the seminal paper on "theory of mind," see David Premack & Guy Woodruff, *Does the Chimpanzee Have a Theory of Mind?*, 1 BEHAV. & BRAIN SCI. 515 (1978).

9 Michael Siegal & Rosemary Varley, *Neural Systems Involved in "Theory of Mind"*, 3 NATURE REVIEWS NEUROSCIENCE 463 (2002). Of course theory of mind predictive mechanisms are not perfect either, and sometimes lead to accidents as well.

10 Ocampo & Kritikos, *supra* note 5, at 263.

11 Mehdi Moussaïd et al., *Experimental Study of the Behavioural Mechanisms Underlying Self-Organization in Human Crowds*, 276 PROC. OF THE ROYAL SOC'Y B: BIOLOGICAL SCI. 2755 (2009).

Autonomous vehicles may challenge this collective ability to avoid harm. Autonomous or "self-driving" cars are computer-controlled vehicles, capable of driving on their own without being operated by a person.[12] In the not too distant future, they are likely to become more common in our physical environment.[13] Due to the safety and efficiency benefits that they are expected to bring, many experts predict that fully autonomous automobiles will be common on the road within the five- to fifteen-year timeframe.[14] As of the writing of this Article, the technology for these self-driving vehicles is quite advanced. Today, experimental autonomous vehicles routinely drive on public roads navigating through traffic, controlled entirely by computer. Collectively, these vehicles have driven close to two million miles completely under their own control.[15]

In many driving contexts, autonomous cars are expected to be safer and more predictable than human drivers.[16] However, in certain scenarios, their movements may be *less predictable*[17] to ordinary people—pedestrians and drivers—who will share their physical space.[18] In these contexts the core theory of mind mechanisms that we rely upon to avoid physical harm may not guide us as accurately when vehicles in our environment are operated, not by other people, but by computer systems.

Consider again the pedestrian at the crosswalk but this time in the context of an approaching *autonomous* vehicle. Imagine that the vehicle slows visibly as it approaches the crosswalk. Using her theory of mind intuition, the pedestrian may believe that the vehicle is communicating its intention to stop by slowing, the way a human driver might.

---

12 LLOYD'S, AUTONOMOUS VEHICLES HANDING OVER CONTROL: OPPORTUNITIES AND RISKS FOR INSURANCE 4 (2014). In fully autonomous vehicles, as discussed *supra*, the role of the person is largely limited to choosing a destination.

13 James E. Young, *How to Manage Robots and People Working Together*, WALL ST. J. (June 2, 2015, 11:10 PM), http://www.wsj.com/articles/how-to-manage-robots-and-people-working-together-1433301051.

14 RICHARD WALLACE & GARY SILBERG, KPMG & CTR. FOR AUTO. RESEARCH, SELF-DRIVING CARS: THE NEXT REVOLUTION, CENTER FOR AUTOMOTIVE RESEARCH (2012); Andrew Del-Colle, *The 12 Most Important Questions About Self-Driving Cars*, POPULAR MECHANICS (Oct. 8, 2013), http://www.popularmechanics.com/cars/a9541/the-12-most-important-questions-about-self-driving-cars-16016418.

15 *See Google Self-Driving Car Project*, GOOGLE (June 3, 3015), https://plus.google.com/+SelfDrivingCar/posts/iMHEMH9crJb.

16 Myra Blanco et al., *Automated Vehicle Crash Rate Comparison Using Naturalistic Data*, VA. TECH TRANSP. INST. (Jan. 8, 2016), http://www.vtti.vt.edu/featured/?p=422.

17 David Benjamin, *Autonomous Cars: "We Will Have Accidents"*, EE TIMES (Oct. 7, 2015, 1:10 PM), http://www.eetimes.com/document.asp?doc_id=1327929 (describing how self-driving vehicles make unexpected movements).

18 As this Article will explain, this is not to say that autonomous cars are not predictable. To the contrary, their actions are predictable from an engineering perspective. Rather, the activities of autonomous cars may not be predictable to ordinary people who are in their vicinity.

However, the vehicle may not have actually detected the pedestrian and may be slowing for an entirely different reason. For example, the vehicle's computer may have a rule that it automatically slows as it approaches every crosswalk, even when it is not intending to fully stop.[19] By relying upon her theory of mind, the pedestrian may misinterpret the vehicle's slowing as a signal that it will stop, leading to a collision that might not have occurred with a human driver.

This Article argues that autonomous vehicles present novel policy issues with respect to movement predictability in shared spaces. Today people share a physical environment with two types of moving machines: those that are controlled by people but have free range of movement (e.g., automobiles) or those that are controlled by computers but are highly constrained in their range of movement (e.g., elevators). With autonomous vehicles, for the first time, people will be sharing a physical environment with computer-controlled machines that can direct their own activities *and* also have free range of movement.[20] The predictive mechanisms that people rely upon to avoid a great deal of physical harm may be challenged as non-human, autonomous moving machines become more common in our shared physical environment.

Part I of this Article will explain how self-driving vehicles work. It examines what "autonomous" means as applied to technology and some key implications of adopting autonomous vehicles in mainstream transportation. Most importantly, it will explain how autonomous vehicles drive themselves. It is crucial to understand the technology in order to appreciate why autonomous vehicles may be less predictable to ordinary people—such as pedestrians—who will share their physical environment.

Part II of this Article explains how people use internal theory of mind mechanisms to model the minds of others. These mental systems allow us to interpret the communicative signals of attention and intention of those around us, predict their movements, and routinely avoid accidental harms. Such internal harm-avoidance mechanisms may be less effective as computer-controlled vehicles enter our physical environment.

In part, the diminishment in predictability occurs because our cognitive systems evolved to predict human behavior and not

---

[19] *See* Chris Urmson, *The View from the Front Seat of the Google Self-Driving Car*, BACKCHANNEL (May 11, 2015), https://medium.com/backchannel/the-view-from-the-front-seat-of-the-google-self-driving-car-46fc9f3e6088 (discussing a similar type of a general computer rule in which a self-driving car will pause slightly at every intersection after a light turns green, regardless of context).

[20] *See* M. Ryan Calo, *People Can Be So Fake: A New Dimension to Privacy and Technology Scholarship*, 114 PENN. ST. L. REV. 809, 814 (2010) (discussing how some experts predict an increase in robots in the consumer space).

computer-controlled activity. However, another contributing factor is the "technological opacity" of autonomous vehicles.[21] A system is "technologically opaque" if it is difficult for an ordinary person to understand what is going on inside that system.[22] Autonomous vehicles are composed of electronics, software, sensors, and mechanical parts. Simply by observing such a machine, a person will not intuitively know where the machine will move next. Such a decision is not externally transparent because it is conducted internally through computer analysis of the vehicle's sensors. Thus, a pedestrian at a crosswalk may not know whether an approaching autonomous vehicle will stop (or even if it has detected her presence) unless the machine has been specifically designed to communicate such information. In general, we can only understand what is going on inside a technological system to the extent that engineers have expressly designed it to communicate that relevant information externally. Engineering design has thus become crucial in the context of autonomous vehicles and safety.

Although this Article repeatedly uses the pedestrian scenario as an example of autonomous vehicle unpredictability, this Part emphasizes it *is not just* about pedestrian conflicts. Rather, that scenario is just one instance of a larger group of predictability conflicts between self-driving cars and people in near proximity, including drivers, cyclists, and passengers.[23]

Part III explores the role of the legal system in mitigating these risks. Fortunately, once the problem has been recognized, it is possible to make the movements of autonomous vehicles more outwardly predictable through technological design choices. Autonomous cars, for instance, can be designed to more clearly communicate to those around them their intentions as they approach high-conflict zones. The government may have varying degrees of involvement to make driverless vehicles more predictable: from fostering more communicative vehicle designs, to standardizing autonomous actions across manufacturers in common driving scenarios, to educating the public about the technology.

Self-driving cars are not the only autonomous systems that may have safety and movement-predictability issues. Experts expect that in

---

21 "Technological opacity" is new terminology used in this Article.

22 *See* Steve Crowe, *A Cyclist's Encounter with an Indecisive Google Self-Driving Car*, ROBOTICS TRENDS (Aug. 26, 2015), http://www.roboticstrends.com/article/a_cyclists_encounter_with_an_indecisive_google_self_driving_car (describing how a cyclist encountered an experimental autonomous vehicle and was unable to determine what it was going to do next).

23 Jon Fingas, *Google Self-Driving Car Crashes into a Bus (Update: Statement)*, ENGADGET (Feb. 29, 2016), http://www.engadget.com/2016/02/29/google-self-driving-car-accident (describing how human bus driver misinterpreted self-driving car's driving intentions, leading to a car accident).

coming decades, other types of autonomous systems, such as robots or airborne drones, will increasingly emerge from specialist contexts (such as factories or laboratories) and into consumer spaces.[24] These systems raise similar concerns. As they begin to move on their own near lay people, it is important that they too be designed to reliably communicate their movements.[25] For instance, if a worker is standing behind an autonomous robot, it is crucial for her to know when the robot is about to take a potentially dangerous action, such as moving backwards, and more generally whether it has detected her presence.[26] The issues raised thus generalize beyond self-driving vehicles to other types of autonomous moving systems.

To be clear, this Article *is not* suggesting that autonomous vehicles are less safe than today's human-operated cars. Quite the opposite is likely true. Most experts predict that autonomous cars will be much safer than human drivers.[27] Nearly ninety percent of automobile accidents are caused by human error.[28] Human drivers are prone to impairments including intoxication, sleep deprivation, and distraction, to which autonomous vehicles are not susceptible.[29] Nor is this Article suggesting that autonomous cars are inherently *unpredictable*. The behavior of autonomous cars is quite predictable to *engineers* who designed them, and autonomous cars quite reliably follow the instructions that they are given the way that most computers do. Rather, this Article is suggesting that we must focus upon making the movements of autonomous cars predictable relative to the intuitions and expectations of *ordinary people* that they will interact with in their physical environment—pedestrians, passengers in the autonomous cars, and other drivers on the road.

---

24 Vinod Baya & Lamont Wood, *Service Robots: The Next Big Productivity Platform*, PWC, http://www.pwc.com/us/en/technology-forecast/2015/robotics/features/service-robots-big-productivity-platform.jhtml?utm_content=buffer10fb2 (last visited July 24, 2016).

25 M. Ryan Calo, *Open Robotics*, 70 MD. L. REV. 571 (2011).

26 Rony Novianto, Benjamin Johnston & Mary-Anne Williams, *Habituation and Sensitisation Learning in ASMO Cognitive Architecture*, *in* SOCIAL ROBOTICS: 5TH INTERNATIONAL CONFERENCE, ICSR 2013, BRISTOL, UK, OCTOBER 27–29, 2013 PROCEEDINGS 249 (Guido Herrmann et al. eds., 2013), http://link.springer.com/chapter/10.1007/978-3-319-02675-6_25.

27 *See, e.g.*, LLOYD'S, *supra* note 12; Del-Colle, *supra* note 14.

28 NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T. OF TRANSP., NATIONAL MOTOR VEHICLE CRASH CAUSATION SURVEY 24 (July 2008); Bryant Walker Smith, *Human Error as a Cause of Vehicle Crashes*, CTR. FOR INTERNET & SOC'Y (Dec. 18, 2013, 3:15 PM) http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes.

29 JAMES M. ANDERSON ET AL., RAND CORP., AUTONOMOUS VEHICLE TECHNOLOGY: A GUIDE FOR POLICYMAKERS, at xv (2016).

I. HOW AUTONOMOUS VEHICLES WORK

This Part will explain how autonomous vehicles work, in a manner accessible to non-technical audiences.[30] The first Section will give an overview as to why self-driving cars present novel problems of predictability. The next Section will explore what it means for a computer system to be "autonomous" and discuss the differences between fully and partially or "semi" autonomous vehicles. The final Section will explore the electronics and software that autonomous vehicles rely upon to drive themselves. That Section will also explain how self-driving cars actually use the technological hardware and software to self-operate on open roads.

A. *Overview: Computer-Controlled Unrestricted Movement*

Fully autonomous or "self-driving" automobiles are vehicles, which "can drive themselves without human supervision or input."[31] The basic contention is that autonomous vehicles—because they are controlled by computers, and are therefore not amenable to our internal introspection capabilities—may be less instinctively predictable to the ordinary people (such as other drivers, cyclists, or pedestrians) that share their physical space. This represents a distinct change from our relationship with the machines in our physical environment today, whose movements tend to be broadly predictable.[32]

On one side, we share an environment filled with moving people and machines operated by people (e.g., automobiles or construction equipment). Machines operated by humans have a considerable range of freedom in where and when they move. Thus, one might expect accidental collisions to be relatively common. However, because such movement decisions are ultimately made by other people, our shared theory of mind mechanisms allow us to reliably signal and predict the movements of those around us.[33] For instance, humans have developed the remarkable ability to walk through dense crowds of randomly moving pedestrians, such as a crowded city sidewalk or a concert venue,

30 Note that organizations developing autonomous vehicles use a number of different technologies and strategies. Thus, the technologies covered many not exist in all approaches to developing autonomous vehicles.

31 *See* LLOYD'S, *supra* note 12, at 4.

32 For instance, today most industrial robots perform fairly repetitive actions and tend to be confined to restrictive spaces. *See* Baya & Wood, *supra* note 24.

33 To a lesser extent there are animals, which are somewhat less predictable.

without physically colliding with one another.[34] We are able to intuitively read the body movements of others to avoid running into them.[35] Physical collisions are relatively rare compared to the total number of interactions.

On the other side, we share a physical environment with automated moving machines whose activities are controlled entirely by computers, such as elevators, escalators, or factory machines. However, the movements of such automated machines are also broadly predictable because these machines tend to be constrained to a limited range of movement. For instance, elevators and escalators move along tracks and highly restricted routes, and factory machines perform repetitive movements in well-defined, often protected, locations.

Autonomous vehicles occupy a middle ground that has little or no comparator today among moving entities. On the one hand, their automated movements are not limited to highly circumscribed, repetitive routes, as are elevators. Rather, autonomous vehicles are capable of driving on ordinary roads, going nearly anywhere a human driver might go. On the other hand, their movement choices are made by computer systems, not by humans. Their movements are, therefore, not intuitively revealed through cognitive introspection and projection.

Unrestricted, computer-directed movement is a novel phenomenon that is likely to challenge certain basic assumptions embedded in our existing legal structure. For instance, tort law (and other areas of law concerned with accidental harm) operates within an overall framework that implicitly presumes that the movements of others will be broadly predictable.[36] More generally, it is likely that a great deal of societal harm is avoided not through the explicit legal rules, sanctions, or incentives, but rather, it is avoided implicitly through the self-preservation activities that people undertake after anticipating what others nearby will do.[37] Accidents that are avoided due to human communication, observation, and prediction, never enter the legal system.

Autonomous vehicles represent a novel and potentially difficult-to-predict class of computer-controlled systems (including robots and drones) in which the machine itself decides what movements to take. To better comprehend these points, it is critical to have an understanding of the underlying technology that allows autonomous vehicles to drive themselves.

---

34 *See* SAAD ALI ET AL., MODELING, SIMULATION AND VISUAL ANALYSIS OF CROWDS: A MULTIDISCIPLINARY PERSPECTIVE (2013).

35 *See* Kendon, *supra* note 3, at 18.

36 *See infra* Part III.

37 F. Patrick Hubbard, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803, 1819 (2014).

B. *Autonomy in Self-Driving Vehicles*

1. What Does Autonomous Mean?

What does it mean to say that a self-driving car is "autonomous?"[38] In the technological context, engineers apply the term "autonomous" to computer controlled systems that make important choices about *their own actions* with little or no human intervention.[39] Autonomous systems are thus able to direct their own activities in the face of an unpredictable or changing physical or data environments.[40] In many cases, these are choices that would otherwise be made by a person in a non-autonomous (human-directed) system.[41]

A simple example of a moving autonomous system is the Roomba vacuum. A Roomba is a small, wheeled household robot that is capable of vacuuming a room entirely on its own.[42] Such a system is considered autonomous because it is the robot itself (via its onboard computers and sensors) that decides where to move and how to avoid obstacles, such as tables and chairs, without being directed by a person.[43] By contrast, a traditional vacuum cleaner is non-autonomous because it is a person, not a computer, that manually directs it around the room and obstacles.

The term "autonomous" can also apply to non-moving computer systems, such as algorithmic financial trading systems.[44] In traditional financial trading systems, a person such as a trader is in charge of deciding what financial securities (e.g., equities, bonds, options) to buy or sell and at what price. By contrast, in autonomous algorithmic financial trading systems, the computer system itself decides which financial instruments to buy and sell, and when, based upon automated

---

38 In general, the word "autonomous" means independent and not subject to outside control. *See Autonomous*, DICTIONARY.COM, http://dictionary.reference.com/browse/autonomous (last visited July 17, 2016); *Autonomous*, MERRIAM-WEBSTER DICTIONARY, http://www.merriam-webster.com/dictionary/autonomous (last visited July 17, 2016).

39 Bruce T. Clough, *Metrics, Schmetrics! How The Heck Do You Determine a UAV's Autonomy Anyway?*, AIR FORCE RES. LIBR. (Aug. 2002), http://www.dtic.mil/dtic/tr/fulltext/u2/a515926.pdf.

40 Jeffrey O. Kephart & David M. Chess, *The Vision of Autonomic Computing*, 36 COMPUTER 41, 42 (2003) ("The essence of autonomic computing systems is self-management.").

41 *See, e.g.*, Bryant Walker Smith, *Automated Vehicles Are Probably Legal in the United States*, 1 TEX. A&M L. REV. 411, 419 (2014) (describing an automated car as "computer direction of a vehicle's steering, braking, and accelerating without real-time human input").

42 Chris Woodford, *Roomba® Robot Vacuum Cleaners*, EXPLAIN THAT STUFF!, http://www.explainthatstuff.com/how-roomba-works.html (last updated Jan. 29, 2016).

43 *Id.*

44 *See* Austen Hufford, *Algorithmic Trading: The Play-at-Home Version*, WALL ST. J. (Aug. 9, 2015, 8:51 PM), http://www.wsj.com/articles/an-algo-and-a-dream-for-day-traders-1439160100.

analysis of data. The system independently executes these purchases or sales without human intervention[45].

In sum, the key characteristic of an autonomous system is that *the system itself* is capable of making decisions about some (or all) of the system's most important activities, with little or no human intervention.[46] In the case of the Roomba robot, the core activities included where to move in the room to vacuum, and how to avoid obstacles; in the case of the algorithmic trading system, these included what financial securities to buy and sell, and when.[47] In the context of autonomous vehicles, important activities under the control of the car itself include steering, accelerating, braking, lane positioning, routing, and following traffic rules and signals.[48]

2. Full Autonomy vs. Semi-Autonomy

It is important to distinguish between fully autonomous and partially autonomous vehicles, as this Article's focus is on fully autonomous vehicles. Engineers classify systems along a spectrum of autonomy, depending upon the extent to which the system makes decisions about its own actions. On one end of the spectrum, if a human is making all of the most important decisions, that system has little or no autonomy. On the other end, if a computer is making all of the most important decisions, there is full autonomy. In the middle, there are partially or "semi" autonomous systems, in which some important actions are decided by humans, and others by computer.

When individuals use the phrases "self-driving," "driverless," or "autonomous" vehicles, they most commonly mean "*fully autonomous*" vehicles.[49] A *fully autonomous* vehicle is one that is capable of driving from one location to another completely on its own, without any human

---

45 This example illustrates that systems do not have to produce movement in the physical world (like autonomous cars or Roombas) to be considered autonomous systems. Rather, computer systems can be autonomous if they interact with information in a self-directed manner without necessarily producing actions in the physical world. ALAN WINFIELD, ROBOTICS: A VERY SHORT INTRODUCTION 12 (2012).

46 The term "autonomous" can be contrasted with a similar term "automatic." Automatic typically refers to the fact that a computer is following precisely a set of preprogrammed instructions on its own. However, automatic systems do not necessarily engage in significant decision-making as to the actions the system is taking. Autonomous, on the other hand, suggests a degree of independent decision-making on the part of the computer. Thus, not all systems that are automatic are autonomous, if the automatic system is not making significant decisions on its own. *Id.*

47 *Id.*

48 NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., PRELIMINARY STATEMENT OF POLICY CONCERNING AUTOMATED VEHICLES (2013) [hereinafter NHTSA].

49 The term "self-driving car" thus refers to *fully* autonomous vehicles.

intervention.[50] If a vehicle is fully autonomous, the vehicle itself makes all major driving decisions—including steering, braking, speed, distance between vehicles, lane-choice, following traffic rules, routing, avoiding obstacles—and the role of the person is limited primarily to choosing the destination.[51] As of the writing of this Article, fully autonomous vehicles are not for sale in the consumer context. Those that exist operate as prototype research vehicles or in controlled commercial settings such as in remote mining areas.[52]

By contrast, partially autonomous vehicles exhibit a mix of human and computer control, with some important activities directed by computer (e.g., emergency braking) and others by humans (e.g., ordinary braking, steering, accelerating).[53] We can, therefore, classify most existing vehicles along the autonomy spectrum, depending upon the degree to which important driving activities are under the control of the human driver instead of a computer.[54]

Most vehicles on the road today exhibit some limited, partial autonomy as some important driving functions of ordinary consumer cars have already been automated.[55] Modern automobiles are designed so that driving functions are separated into distinct "subsystems" that control specialized driving activities.[56] For instance, a vehicle might have one subsystem for steering and another for braking; these subsystems work together to allow overall driving functionality.[57] Today, most consumer cars contain some subsystems that are autonomous in the sense that a computer, rather than a person, initiates

50 LLOYD'S, *supra* note 12, at 7. As discussed, the role of the person is limited to choosing a destination.

51 Smith, *supra* note 41.

52 *See* discussion *supra* Section I.C.

53 NIDHI KALRA, JAMES ANDERSON & MARTIN WACHS, RAND CORP., LIABILITY AND REGULATION OF AUTONOMOUS VEHICLE TECHNOLOGIES 3 (2009) ("A widely used approach to categorizing these technologies is by the degree to which they intervene in the driving of the vehicle.").

54 The National Highway Traffic Safety Administration has created a classification system with five levels of autonomy for characterizing the level of autonomy in vehicles. "Level 0" involves no autonomy, and the driver is in control of all driving functions. In a "Level 1" vehicle, the vehicle will take over limited individual driving functions, such as automated vehicle stability. "Level 2" involves more significant automation, where the vehicle combines two or more driving functions. An example is adaptive cruise control and lane keeping that would allow a driver to remove his hands off the wheel for limited periods of time. In "Level 3" vehicles, nearly all driving activities are automated, but a human driver is required to take over in case of an emergency. "Level 4" is full autonomy, where the entire trip is automated, and the driver has no functional role. *See* NHTSA, *supra* note 48 , at 3–6.

55 *Id.*

56 Wuhong Wang et al., *A Framework for Function Allocations in Intelligent Driver Interface Design for Comfort and Safety*, 3 INT'L J. OF COMPUTATIONAL INTELLIGENCE SYSTEMS 531 (2010).

57 *Id.*

important driving activities.[58] For instance, air-bag subsystems automatically deploy when crash sensors detect a collision; humans do not control when airbags deploy.[59] Similarly, anti-lock brakes or traction control systems typically engage on their own depending upon automated detection of low-traction conditions. Thus, some limited self-directed driving activity is already familiar today.[60]

However, a more significant trend in semi-autonomous driving can be found in Advanced Driver-Assistance Systems (ADAS). ADAS refers to a series of emerging technologies that automatically take control of particular driving functions. ADAS systems have been available since about 2012 as optional features and are becoming common in ordinary consumer vehicles.[61] For instance, consumer automobiles increasingly feature automatic emergency braking systems that autonomously brake under particular circumstances.[62] These systems detect when a vehicle is about to collide with an object and will automatically initiate the brakes to prevent or mitigate the collision using sensors.[63] Other emerging ADAS features include lane-keeping systems (that automatically correct steering to keep a driver within lane boundaries), automatic parking, and adaptive cruise control systems (that automatically accelerate, brake, and maintain a safe distance behind another vehicle on the highway by detecting distances and adjusting speed).[64] Notably, even with ADAS systems, human drivers still retain control over the vast majority of driving functions.[65] Thus, even though they substantially increase the degree of driving automation, vehicles equipped with ADAS features are still considered *partially* autonomous vehicles.

The point is that, given the increasing presence of ADAS systems in consumer cars, the transition to fully autonomous automobiles will be less of a substantial leap over existing consumer technology than is often presumed. Existing ADAS features in consumer vehicles already represent a substantial movement along the spectrum towards more autonomous driving. Thus, although much of the public focus is on the fully autonomous cars of the future, many of the ingredients of self-driving cars are in fact already here. Several of the ADAS technologies in

---

58 KALRA, ANDERSON & WACHS, *supra* note 53, at 1.

59 *Id.*

60 *Id.*

61 *Id.*

62 ANDERSON ET AL., *supra* note 29, at 15.

63 *Id.*

64 RATAN HUDDA, CLINT KELLY & GARRETT LONG ET AL., FUNG INST. SELF DRIVING CARS pt. 2.1, at 3 (2013).

65 Whereas some of these automated systems merely alert the driver as to a dangerous condition (such as blind-spot warning systems), others, such as collision avoidance automatic braking systems, automatically intervene and take over driving functionality when necessary.

ordinary consumer cars today are the same technologies that will be used in future fully autonomous vehicles.[66]

Although ADAS features are important steps on the road to fully autonomous cars, the focus of this Article will not be on such partially autonomous driving. Rather, this Article will explore *fully autonomous cars*, meaning vehicles that are capable of driving from one location to another on their own without any human intervention. From this point on, this Article will therefore use the words "autonomous" and self-driving to mean *fully autonomous cars,* and will specifically note when discussing partially autonomous systems.

C. *State of Autonomous Vehicle Technology*

The technology for fully autonomous vehicles is today quite advanced. Self-driving cars have driven over one million miles on public roads, operated entirely on their own.[67] In these scenarios, a computer is steering the car, accelerating, braking, and following traffic signs or signals. Humans are not the drivers but are instead passengers. Sophisticated systems aboard the self-driving car take in information from sensors that analyze the road and the nearby surroundings to make automated decisions about where and when to drive and stop. While fully autonomous vehicles are not yet available to consumers,[68] they are in commercial use today, typically in limited or remote settings. For instance, driverless trucks are used to transport mining materials in the sparsely populated Outback in Australia, and self-driving tractors are increasingly being used by farmers on agricultural fields.[69]

Self-driving vehicles offer three main benefits over traditional vehicles. First, most experts predict that self-driving cars will be safer drivers than people. Over ninety percent of car accidents today are attributed to human error caused by factors such as intoxication, inattention, sleepiness, or extreme speeding.[70] Self-driving vehicles will

---

66 For instance, some current consumer vehicles allow for significant driving automation in limited settings using adaptive cruise control and other ADAS features. Notably, Tesla released its advanced "Autopilot" mode in 2015, which allows for semi-autonomous highway driving. Katie Fehrenbacher, *How Tesla is Ushering in the Age of the Learning Car*, FORTUNE (Oct. 16, 2015, 12:53 PM), http://fortune.com/2015/10/16/how-tesla-autopilot-learns.

67 David Robson, *The Truth About Driverless Vehicles*, BBC (Oct. 13, 2014), http://www.bbc.com/future/story/20141013-convoys-of-huge-zombie-trucks.

68 Some vehicles, such as Tesla with "Autopilot," exhibit advanced partial autonomy but are not fully autonomous.

69 *See* Robson, *supra* note 67.

70 *See* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., NATIONAL MOTOR VEHICLE CRASH CAUSATION SURVEY (2008); NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., TRAFFIC SAFETY FACTS: ALCOHOL IMPAIRED DRIVING (2014); Naomi Kresge, *Smart Self-Driving Cars Still Need to Factor in Human Error*, BLOOMBERG: TECH. (June 7, 2015, 11:00 PM), http://

not suffer from these issues, and some expect overall accidents to decrease between thirty and eighty percent once self-driving cars are broadly available.[71] Second, self-driving cars offer convenience, as human drivers become passengers who can do other things, such as reading or working, while riding. Finally, autonomous vehicles will offer increased mobility to those who may be unable to drive themselves, such as elderly or disabled populations.[72]

Although the technology is approaching maturity, there are still technological, legal, and social issues that need to be overcome before fully autonomous vehicles are sold to consumers.[73] While the vehicles operate very well on highways and in clear weather conditions, the technology sometimes has difficulty driving in certain conditions such as snow or in dense urban environments.[74] Researchers are still working on the problem of self-driving cars that can handle all weather conditions and environments. Additionally, some of the equipment necessary to create self-driving cars is currently prohibitively expensive for typical consumer purchase.[75] Finally, numerous legal and policy issues need to be resolved at the state and federal level. For instance, one issue currently being debated: should state or federal law require self-driving cars to have licensed drivers in the vehicle even when the cars are driving themselves, and must human drivers be prepared to take control in an emergency?[76]

Because of these legal and technological barriers, there are a wide range of estimates as to when fully autonomous vehicles will be sold to the public. Most experts predict the first consumer sale will occur

---

www.bloomberg.com/news/articles/2015-06-08/cars-smart-enough-to-drive-still-need-to-overcome-human-error.

71 Michele Bertoncello & Dominik Wee, *Ten Ways Autonomous Driving Could Redefine the Automotive World*, MCKINSEY & CO. (2015), http://www.mckinsey.com/insights/automotive_and_assembly/ten_ways_autonomous_driving_could_redefine_the_automotive_world.

72 *See id.*

73 Alex Davies, *Google's Self-Driving Cars Aren't as Good as Humans—Yet*, WIRED (Jan. 12, 2016, 7:46 PM), http://www.wired.com/2016/01/google-autonomous-vehicles-human-intervention (suggesting that current self-driving vehicles are almost, but not quite as good as human drivers as of 2016).

74 Keith Naughton, *Driverless Cars Also Struggle in the Snow*, BLOOMBERG: TECH., (Feb. 9, 2016, 7:01 PM), http://www.bloomberg.com/news/articles/2016-02-10/robot-cars-succumb-to-snow-blindness-as-driving-lanes-disappear.

75 Lidar sensors are currently expensive, ranging between $10,000 and $70,000 just for the sensors alone. *See* HUDDA ET AL., *supra* note 64.

76 *See, e.g.*, Smith, *supra* note 41, at 483. An important debate among self-driving car research areas is the degree to which human drivers should even have the ability to take over driving. Some researchers argue that it is unrealistic, and perhaps dangerous, to have self-driving cars drive themselves, but then require human passengers to suddenly be required to take over. Lee Gomes, *Hidden Obstacles for Google's Self-Driving Cars*, MIT TECH. REV. (Aug. 28, 2014), http://www.technologyreview.com/news/530276/hidden-obstacles-for-googles-self-driving-cars.

somewhere between 2020 and 2035.[77] However, even if consumer-level autonomous vehicles begin to appear by 2020, as some are predicting, substantial penetration of the market is likely to take much longer, with fully autonomous cars not reaching a high proportion of vehicles on the road until the 2030's or later.[78] It is important to emphasize that there is likely to be a lengthy transition period of ten to forty years after the first consumer sale before self-driving cars will abound in any substantial quantity. Even when self-driving cars emerge, they are likely to coexist with ordinary cars for a long period.

D. *Technology of Self-Driving Vehicles*

This Section will explore the technology underlying autonomous driving. Although there are a number of distinct approaches to creating self-driving cars, this Section will highlight the most common strategies.

At a high level, autonomous vehicles use technology to assess three primary questions:

1) Where are they located?

2) What objects are around them?

3) Where is it desirable, legal, and safe to move next?[79]

Within this framework, a self-driving vehicle must be able to understand driving features such as traffic lights, stop signs, and lane markings and navigate within an environment filled with moving objects such as automobiles, pedestrians, cyclists, and animals.

1. Hardware: Sensors

Autonomous vehicles address the three questions above through the use of sensors. A sensor is a technological device that gathers information about the nearby environment (and about the vehicle itself) and relays that information to the vehicle's onboard computers.[80] For instance, many autonomous vehicles use *radar* sensors to detect the location of surrounding objects such as automobiles in another lane.[81] Radar systems determine the position of such objects by emitting radio

---

77 Mitch Turck, *State of Autonomy: July Recap*, MEDIUM (Aug. 2, 2015), https://medium.com/@mitchturck/state-of-autonomy-july-recap-be5bf4dd91e9.

78 *Id.*

79 ANDERSON ET AL., *supra* note 29.

80 *Sensor*, MERRIAM-WEBSTER, http://www.merriam-webster.com/dictionary/sensor (last visited July 17, 2016); WALLACE & SILBERG, *supra* note 14; ANDERSON ET AL., *supra* note 29.

81 *Id.*

waves that naturally reflect off of nearby solid surfaces.[82] If a radio-wave is emitted and then reflected back, this is an indication that some object is there.[83] The radar system can calculate the location, speed, movement, direction of the object by observing the angle, timing, and strength of the reflected wave.[84] Such data can then be relayed to the vehicle's on-board computer system to map the position and movement of nearby automobiles. Radar provides an illustrative example of using a sensor to gather information necessary for autonomous driving.

In addition to radar, autonomous vehicles typically rely upon several other types of sensors, including Global Positioning Satellite (GPS) receivers, sonar, video cameras, inertial navigation systems, and lidar, which is essentially laser-based radar.[85] Collectively, these sensors provide information about the core issues involved in driving including the vehicle's current position (e.g., what street is the vehicle currently on and in which lane?), movement (e.g., what is the vehicle's current speed and direction?), nearby movable obstacles (e.g., are there moving or stopped vehicles, pedestrians, or bicycles nearby?), nearby fixed obstacles (e.g., are there curbs, signs, or buildings in the near vicinity?), and surrounding traffic-safety features (e.g., are there relevant traffic lights, stops signs, or lane markings that need to be observed?).[86] The operation of sensors such as radar, lidar, GPS, will be discussed in more detail later.

2. Annotated Digital Maps

Besides sensors, many vehicles rely upon pre-built digital maps for autonomous driving.[87] Such digital maps contain the expected suite of geographic information, such as the overhead layouts of roads and the associated coordinates (i.e., longitude and latitude) for each point on the road. On one level, these digital maps are broadly similar to what consumers might encounter in widely available vehicle navigation systems.[88] Importantly, however, digital maps for self-driving vehicles

---

82 STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 601, 928 (3d ed. 2014).

83 *Id.*

84 *Id.*

85 Jesse Levinson et al., *Towards Fully Autonomous Driving: Systems and Algorithms*, IV IEEE INTELLIGENT VEHICLES SYMP. 163 (2011), http://cs.stanford.edu/people/teichman/papers/iv2011.pdf; Jesse Levinson, Michael Montemerlo & Sebastian Thrun, *Map-Based Precision Vehicle Localization in Urban Environments*, ROBOTICS: SCIENCE AND SYSTEMS PROCEEDINGS III (2007).

86 ANDERSON ET AL., *supra* note 29.

87 HUDDA ET AL., *supra* note 64, at pt. 2.4, at 4.

88 UMIT OZGUNER, TANKUT ACARMAN & KEITH REDMILL, AUTONOMOUS GROUND VEHICLES 194–96 (2011).

often contain a significant amount of additional information that specifically facilitates autonomous driving.[89]

First, these maps typically have detailed, road-level images of most street locations.[90] These images are typically 360-degree laser scans of roads, taken from the ground-level perspective of a driving vehicle, analogous to what one might encounter on services such as Google Maps Street View.[91] Companies obtain these images by pre-driving the roads that autonomous vehicles will ultimately ride upon. Specialized mapping vehicles equipped with lidar laser scanners (and other sensors) capture and store important visual details for each road portion that will later be used by a self-driving car. Each road location image is precisely affixed with its correct GPS location.[92]

Such pre-collected map images can later be retrieved by an autonomous vehicle as it arrives at each location. The vehicle can examine previously collected images to know what the current road is supposed to look like and what driving features should be there. An autonomous vehicle can also compare its current surroundings to its pre-loaded database of the images to help determine precisely where it is located. Having street-level images of each geographic location, in addition to traditional overhead map coordinates, can significantly improve the effectiveness of autonomous driving.

Second, these digital maps are often manually annotated with information about important driving features such as traffic lights, signs, driveways, and lane markings.[93] Human map specialists meticulously analyze pre-built digital maps and then add crucial driving information, such as a traffic signal or lane path through an intersection, to precise locations on the digital map.[94]

Annotated maps are useful because autonomous vehicles can become aware of important information about a given location as it approaches. For instance, imagine a self-driving car approaches a particular intersection and that a human map specialist has previously manually annotated the map to indicate that there is a traffic signal at this location. As the car arrives, it can examine the pre-annotated digital map and determine that the intersection should have a traffic signal. The vehicle can then double check this pre-loaded information with live sensor information (e.g., using the vehicle's video camera to ensure that

89 Gomes, *supra* note 76.

90 By contrast, most traditional maps tend to have an overhead view of, not a detailed street-level, 360-degree photographs or images of roads.

91 *Google Street View*, GOOGLE MAPS, https://www.google.com/maps/views/streetview?gl=us (last visited July 17, 2016).

92 Levinson et al., *supra* note 85.

93 Gomes, *supra* note 76.

94 Levinson et al., *supra* note 85, at 164–67; David Autor, *Polanyi's Paradox and the Shape of Employment Growth* 33 (Nat'l Bureau of Econ. Research, Working Paper No. 20485, 2014).

there is indeed a traffic signal there), to make much more accurate and safer driving decisions.[95] Combining detailed pre-built mapping with live sensing greatly improves the vehicle's ability to accurately detect traffic signals, stop signs, and other import driving features, compared to relying upon live sensor information alone.

There is a crucial difference between pre-built map information, and the real-time information from the vehicle's sensors. The pre-built maps contain data that was collected at some point in the *past*. By contrast, the vehicle's on-board sensors detect "live" information about the vehicle's immediate surrounding in the current moment. Thus, there is the possibility that there could be a disagreement between information found in a pre-built digital map, and reality, as the road and traffic conditions might change in between the time the map was created and the current moment.[96] For instance, a pre-built digital map created a month earlier might indicate that a particular intersection does not have a traffic light, but as an autonomous vehicle approaches the intersection, its on-board cameras might detect a traffic light that had been installed by the city the day before. In such cases, vehicles are capable of relying solely upon its sensors to safely and accurately navigate in changed conditions.[97] Additionally, it is possible to update a map dynamically based upon reports of changes from multiple confirming vehicles (e.g., 100 vehicles each relay back to a central mapping computer that they have detected a new traffic light at a particular intersection and the map is updated).

In sum, many discussions of self-driving technology focus on sensors, but it is important to emphasize the degree to which self-driving functionality often depends upon pre-built digital maps. Different research strategies rely upon pre-built maps to a greater or lesser degree. In general, when a vehicle can combine past information from pre-built digital maps along with live information from its sensors about its surroundings, this is often the most effective strategy for achieving highly reliable autonomous driving.[98]

---

95 Levinson et al., *supra* note 85, at 167.

96 Vince Bond, Jr., *Up-to-the-Minute Maps Will Be Critical for Autonomous Cars*, AUTOMOTIVE NEWS (Sept. 13, 2014, 12:01 AM), http://www.autonews.com/article/20140913/OEM06/309159962/up-to-the-minute-maps-will-be-critical-for-autonomous-vehicles.

97 In principle, a vehicle might navigate solely based upon the information coming in from its sensors (a sensor-only strategy) without relying upon annotated map. In such a case, the vehicle would rely primarily upon its sensors to detect traffic features such as lane markings, stop signs, or traffic lights, without heavy reliance upon a pre-constructed digital map heavily annotated with driving features.

98 WALLACE & SILBERG, *supra* note 14.

3. Coordinating Computer System

The third crucial technology besides sensors and maps is the coordinating computer system. Such a system organizes and plans all of the vehicle's activities. The system combines data from the sensors and map and uses a variety of sophisticated computer algorithms to determine whether it is safe, useful, and lawful to move the vehicle to a new position. If so, it directs the vehicle to that new position.

E. *Process of Autonomous Driving*

Having surveyed the general technological hardware of self-driving cars, it is important to understand how self-driving cars use this hardware to drive autonomously.

1. Process of Sensing, Planning, and Acting

At a high-level, many autonomous vehicles drive using a three-stage process—sense, plan, act—that is common in robotics generally.[99] In the *sensing* phase, the vehicle uses its multiple on-board sensors—radar, lidar, GPS, cameras—to gather information about where the car is located and what is around it. In the *planning* stage, information from multiple sensors is fed to the coordinating computer system, which analyzes this data. The computer creates a digital representation of nearby objects and driving features based upon sensor information. It then integrates this information into the overall plan as to where the vehicle is attempting to go. Using complex software, the on-board computer then makes a determination about where it is safe and legal to move next (e.g., there is no object ahead, so it is safe to move forward ten feet). Finally, in the *acting* phase, the on-board computer actually moves (or stops) the vehicle in a manner that is consistent with the computer plan (e.g., drive the vehicle ten feet forward in the lane). The computer moves the vehicle by electronically activating the appropriate driving systems such as the accelerator, brakes, or steering.

Importantly, this "sense-plan-act" process of gathering information about the vehicle's nearby environment and analyzing it is a continuous cycle, which happens repeatedly, hundreds, or thousands of time per second.[100] This allows the autonomous vehicle to adapt to sudden

---

99 ANDERSON ET AL., *supra* note 29, at 58.

100 *See* Baya & Wood, *supra* note 24 (discussing simultaneous location and mapping (SLAM) and constant updating in robot architectures).

changes in the physical environment. For instance, a vehicle might, after scanning, determine that it is safe and desirable to change lanes. However, a moment later, a bicycle that was not previously there might enter the parallel lane. Because the scanning-planning-acting process is continual, the vehicle's sensors can rapidly detect most relevant changes in conditions and adjust. The on-board computer can, on the basis of this newly sensed information, update the plan according to the changed circumstances and cancel the lane change. It is the fact that the process of scanning and adapting is continual that helps the automobile to avoid dangerous situations.

In general, autonomous vehicles must "perceive" and understand their surrounding physical environment. During this sensing phase, the vehicle has to make three primary determinations: where is the vehicle located—both generally (e.g., on Route 36 heading east) and specifically (e.g., what lane on Route 36)—what objects and obstacles are around it (e.g., other vehicles or pedestrians) and what are the major driving features (e.g., lane markings, stop signs, intersections, traffic signals) that it must be aware of to drive safely and legally.[101] The next Subpart will explore these processes in detail.

2. Location: Where is the Vehicle Located?

In most cases, self-driving vehicles operate best when they can precisely determine where they are currently located. This process of geographic location determination is known as "localization."[102] To determine their current location, self-driving vehicles often use a two-stage system: 1) they first use GPS to gain a rough approximation of their location, detecting the current road and direction, and 2) then more precisely determine physical placement on the detected road within a few centimeters of actual location (e.g., which lane, and position within lane) using data from other sensors.

GPS can provide a good first approximation as to where an autonomous vehicle is located. GPS operates by using satellites that broadcast precisely timed signals from known positions in space. Self-driving cars use GPS receivers to examine the timing of signals from each satellite to calculate the latitude and longitude of the vehicle. Readers may be familiar with GPS mapping as it is essentially the same technology used in navigation systems that are common on consumer automobiles and smartphones today.

---

101 LLOYD'S, *supra* note 12, at 6.

102 Jesse Sol Levinson, Automatic Laser Calibration, Mapping, and Localization for Autonomous Vehicles (Aug. 2011) (unpublished Ph.D. dissertation, Stanford University, https://stacks.stanford.edu/file/druid:zx701jr9713/JesseThesisFinal2-augmented.pdf).

While GPS technology is sufficient for many mapping purposes (e.g., road and compass direction), it is not precise enough for the task of autonomous driving. Autonomous cars must be able to determine their location down to a precision of several centimeters. An error of position by as little as twenty centimeters could accidentally place the vehicle into the oncoming lane. Problematically, GPS can be inaccurate by as much as five meters due to interference and other limitations.[103] Therefore, autonomous cars typically must supplement the approximate GPS location with more precise means of determining road positioning.[104]

To more accurately determine road positioning, autonomous vehicles often supplement GPS data with information from lidar—"Light Detection and Ranging"—sensors. As described earlier, radar systems detect nearby objects by analyzing the time it takes for *radio waves* to travel to an object and reflect back. Lidar systems are analogous to radar, except that lidar reflects *laser beams* off of nearby objects to detect their location.[105] Lidar systems can calculate the speed, position, and distance of nearby objects by measuring how long it takes laser beams to reflect back.[106]

An important advantage of using lidar is its precision. Since lidar uses laser beams that are smaller than radio waves to make measurements, it can accurately determine the distance to nearby driving features within millimeters in dry conditions. For instance, lidar can be used to detect the precise distance to reflective surfaces such as white lane boundary lines painted on road surfaces.

Lidar's ability to accurately detect the distance to road features, such as nearby lane boundaries, assists the vehicle in determining its precise location.[107] Recall that autonomous vehicles have access to digital maps that were created by pre-driving the roads, and that these maps often have detailed images of each road location, including lane markings. Once an autonomous car determines its approximate location using GPS, it can then retrieve a pre-collected image of what the road is supposed to look like in that general area. It can then compare its live lidar scans of the surrounding area against the pre-built digital images to estimate, probabilistically, where it likely is in the map. For instance, the autonomous vehicle might compare live information from its lidar sensor about detected painted lane markings to previously collected

---

103 Levinson et al., *supra* note 85.

104 *Id.* at 164.

105 *Id.*

106 Ryan Whitwam, *How Google's Self-Driving Cars Detect and Avoid Obstacles*, EXTREMETECH (Sept. 8, 2014, 3:45 PM), http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles.

107 Levinson et al., *supra* note 85, at 164.

information from digital maps with the known location of lane markings, to make a determination as to which lane it is in.[108] Using these measurements and comparing it with the annotated map, it can then precisely determine its location on a given road (e.g., which lane) within centimeters.[109]

To avoid confusion, it is important to note that in many autonomous vehicles, lidar is used to perform two distinct functions: 1) determining the precise road location of the autonomous vehicle, and 2) detecting objects surrounding the vehicles such as other automobiles. The previous Section discussed the precision-location use of lidar, and the subsequent Section will discuss how lidar is also used to detect surrounding objects.

3. Sensing: What Obstacles Must Be Avoided?

The most critical part of autonomous driving is avoiding obstacles, such as surrounding automobiles, pedestrians, curbs, and bicycles. To do so, autonomous vehicles must detect and determine the location of such obstacles. If the objects are moving—such as parallel cars—it must determine their current speed and direction. However, determining the current location of moving objects is not enough; autonomous cars must also be able to predict their future location based upon their current speed and trajectory.

For instance, imagine that an autonomous vehicle detects a bicycle ahead and moving perpendicular to the path of moving vehicle. Because the autonomous vehicle and the surrounding objects are continually changing location, the autonomous vehicle must be able to predict where the bicycle will be a moment from now by analyzing the bicycle's current speed and heading relative to the autonomous vehicle's future position. It needs to estimate where the bicycle will be to ensure that the vehicle's current and planned movement actions are safe (e.g., is the bicycle's current speed and heading likely to put it in the path of the autonomous vehicle's currently chosen path?). To determine the location of moving and fixed objects, autonomous vehicles use a mix of the sensors previously mentioned—most typically lidar, radar, and video cameras.

a. Lidar for Obstacle Detection

Lidar has come to play an important role in allowing autonomous vehicles to detect the obstacles around them. Lidar systems are typically

---

108 *Id.*
109 *Id.*

mounted on the roof of the autonomous vehicle and rapidly rotate 360-degrees.[110] This high placement and rapid rotation allows lidar to detect objects on all sides of the vehicle, including those behind the vehicle at a rate of up to a million readings per second. Lidar systems, because they use lasers, are precise in their location determinations of objects, on the order of millimeters in ideal conditions. Thus, a lidar system can reliably discern the distance of a tiny object up to 100 meters away.[111]

Autonomous vehicles frequently use lidar to create live internal computer representations of the moving and stationary objects—such as nearby cars—currently around the vehicle. This is sometimes confusingly called the lidar "live map." Such a lidar "live map" is created in real-time of the vehicle's immediate surroundings and should be distinguished from the pre-built digital maps that have been created at some point in the past and which cover a much broader geographic area. These real-time lidar live maps are crucial to predicting the future behavior of nearby vehicles and ensuring that the autonomous vehicle will not collide with other objects.[112]

b. Radar

In addition to lidar, many autonomous vehicles use radar to detect the position and speed of surrounding objects. Radar has a few advantages over lidar in certain positioning tasks. For one, the range of radar is much greater, up to several hundred meters or more.[113] Moreover, radar systems tend to be much less expensive than lidar.[114] The most important advantage is that radar is useful for assessing the speed of multiple moving objects, such as nearby vehicles, in real-time.[115] The primary disadvantage of radar compared to lidar, is its precision, which can be off by several inches to feet in detecting the location of stationary obstacles.[116] For this reason, autonomous vehicles

110 ANDERSON ET AL., *supra* note 29, at 61.

111 Whitwam, *supra* note 106.

112 The major advantage of lidar, over other sensors, is its precision. However, lidar has a few disadvantages worth mentioning. For one, as of the writing today, lidar systems tend to be quite expensive (although prices are expected to fall with mass demand). However, a more important limitation is the range. The maximum range of a lidar system tends to be about 100 meters.

113 Whitwam, *supra* note 106.

114 Matt McFarland, *The $75,000 Problem for Self-Driving Cars is Going Away*, WASH. POST (Dec. 4, 2015), https://www.washingtonpost.com/news/innovations/wp/2015/12/04/the-75000-problem-for-self-driving-cars-is-going-away (describing how, as of the writing of this Article, lidar systems tend to cost close to $75,000; but they may be become less expensive in the future).

115 Whitwam, *supra* note 106.

116 As suggested, this technology is similar to that used by fully autonomous vehicles to detect nearby obstacles. KALRA, ANDERSON & WACHS, *supra* note 53.

often use information from radar and lidar in parallel to gain different sources of information about the location of obstacles.

c. Video Cameras

Finally, many autonomous vehicles use video cameras to detect the location and speed of nearby obstacles. A typical arrangement involves two or more video cameras spaced around the vehicle at known distances. Spacing multiple video cameras in this way allows the vehicle's computer to receive parallel images of the same objects but from slightly different angles. Viewing the same object from multiple known angles allows the computer to estimate an object's distance, a phenomenon known as stereopsis.[117] This use of parallel video cameras is similar to the way that people visually assess distance. The human brain uses slightly different images of the same objects from the left and right eyes to estimate depth and distance.

Besides detecting the *position* of objects, video cameras can be a valuable source of information about other features that are crucial to driving.[118] For instance, video cameras are often used to read words on traffic signs or determine whether a traffic signal is green or red. In using video camera data in this way, autonomous vehicles employ techniques from the field of "machine vision," the field that studies algorithmic approaches to making sense of visuals, such as discerning that a particular object is a stop sign, or identifying a traffic signal and its current color.[119] Because visual cues play an important role in driving, cameras can capture information, such as color or language, that the other sensors, such as lidar or radar, may not be well-suited to retrieve. In sum, in the sensing phase, the self-driving vehicle uses sensors such as lidar, radar, and cameras to gather important information about the vehicle's surroundings, such as location, nearby obstacles, and traffic features.

4. Planning: Where Is it Safe, Legal, and Desirable to Move?

The next phase in autonomous driving is "planning," where the vehicle determines where it is safe, legal, and desirable to move next.

---

117 RUSSELL & NORVIG, *supra* note 82, at 949–50.

118 Video cameras are used to identify and classify the types of objects around the vehicle. For instance, is a nearby object a bicycle, a pedestrian, or a vehicle?

119 It is important to note that some approaches to autonomous driving rely more heavily upon certain sensors than others. For instance, some self-driving approaches rely more heavily upon video cameras for obstacle detection whereas other approaches depend heavily upon lidar and use comparatively little video and machine vision to drive. Still, other self-driving approaches use multiple sensors and aggregate the collective input from lidar, radar, and video camera sensors to detect obstacles.

During this phase, the vehicle's supervising computer system takes all of the data from its various sensors and uses it to coordinate the vehicle's future actions. Combining live sensor information with static information from the pre-annotated digital maps, the computer uses a variety of sophisticated computer algorithms to plan the vehicle's next movement: steering, acceleration, braking, and resting. In many ways, planning is the heart of autonomous driving, as the computer system is making automated decisions about how to direct its own actions based upon what it has detected about its surroundings.

Typically, the vehicle's central computer uses sensor information to build an internal live representation of all of the objects immediately around the vehicle (e.g., other automobiles and pedestrians), their position and speed, and computes their predicted positions. The computer also uses information from the sensors and the digital map to identify the position of fixed obstacles such as curbs and important driving features such as lane markings, signs, and traffic signals. As will be discussed, the vehicle must be able to identify different types of moving objects, distinguishing a bicycle rider, from a pedestrian, from a motorcycle in order to respond sensibly. Finally, the vehicle must also be able to respond to core and dynamic traffic features, such as a traffic signal changing from red to green. All of this information is fed into a series of planning algorithms that take into account driving paths, traffic laws, driving conventions, safety, and comfort, to create a way forward.

a. Machine Learning

Driving is such a complicated and unpredictable task that it is difficult to program this activity using a series of prewritten computer rules that instruct when to brake, accelerate, or steer.[120] For this reason, many autonomous vehicles instead rely upon a flexible programming technique known as "machine learning." Generally speaking, machine learning refers to computer algorithms that are able to automatically "learn" or improve in performance on some task over time, such as driving.[121] Such algorithms learn how to take action by analyzing data and detecting patterns in that data that are informative of the task at hand. Often, one "trains" a machine learning algorithm to be better at some task by providing it with relevant good (and bad) examples. The machine learning algorithm can essentially program itself by finding patterns among the examples that lead to good (or bad) outcomes. Thus, in machine learning, loosely speaking, the computer learns the

120 Susan Kuchinskas, *Crash Course: Training the Brain of a Driverless Car*, SCI. AM. (Apr. 11, 2013), http://www.scientificamerican.com/article/autonomous-driverless-car-brain.

121 Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87 (2014).

"rules"[122] to guide its actions on its own, rather than having those rules pre-programmed by human programmers.

An example from a different context will illustrate the point of learning from examples. Machine learning algorithms are often used to automatically detect spam emails. However, an important point is that such machine learning spam algorithms are not explicitly programmed by computer programmers with a set of rules that allow them to distinguish spam from wanted emails. Rather, such algorithms are designed to learn this information on their own by analyzing spam and non-spam emails and detecting the telltale patterns of spam. After observing many examples of spam and wanted emails, such algorithms are able to automatically identify patterns of relevant indicia of spam. For instance, a machine learning algorithm might automatically discern that phrases such as "Earn Cash" are statistically more likely to occur in spam than in wanted emails.[123] It can use patterns like this that it has automatically detected to make reliable predictions about whether incoming emails are or are not spam.

Analogously, self-driving vehicles do not primarily drive themselves based upon a series of pre-programmed computer rules about when and where to steer, accelerate, or brake.[124] Rather, such systems typically use machine learning algorithms that have been "trained" to drive by analyzing examples of safe driving, and automatically generalizing about the core patterns that constitute effective driving from these examples. For instance, one approach is to have a person drive a vehicle on the open roads with a machine learning algorithm observing the human driver's actions (and data about those actions) and automatically generalizing about proper driving approaches. In such a case, the machine learning algorithm can observe a human driver's steering, braking, and acceleration data in various locations, and with various sensor readings of surrounding vehicles, and analyze this data for patterns. After analysis, the algorithm might detect, for instance, that braking always occurs when there is a stopped vehicle detected in front.[125] It can then "learn" an association, on its own, such as that it should generally stop when it detects a stopped vehicle in front.

---

122 *Id.* at 94. It is important to note that I am using the word "rules" loosely, speaking for explanatory purposes. In reality, many machine learning algorithms do not formulate explicit formal rules for conduct, but rather, encode their behavior in non-rule models.

123 *Id.*

124 To be clear, complex systems often involve a mix of explicit rules and learning algorithms. So many autonomous driving systems do have a mixture of explicit, general rules (e.g., always pause at an intersection before proceeding) and machine learning algorithms.

125 Andrew Ng, *Lecture 57—Autonomous Driving*, COURSERA, https://www.coursera.org/learn/machine-learning/lecture/zYS8T/autonomous-driving (last visited July 20, 2016) (video discussing training machine learning algorithm to drive vehicle).

Similarly, the algorithm might also detect that human drivers generally steer slightly away when approaching lane boundaries. It can then learn from this data a general pattern that it should stay within lane boundaries. With many such detected correlations between sensor input and human driving actions, the machine learning algorithms can develop complicated models that ultimately give it the ability to steer and navigate the road on its own in novel settings, based upon analyzing live sensor data about the surrounding road conditions.[126] Thus, these algorithms can learn to detect patterns in the data generated by human drivers that are associated with proper driving, and can effectively "learn," on its own, the appropriate activities that constitute safe driving.[127]

There are other tasks besides movement that machine learning is used for in the autonomous vehicle context. Machine learning is also used to help self-driving vehicles identify the different types of objects around it, for instance, distinguishing bicycles from, automobiles, motorcycles, cyclists, or pedestrians.[128] Properly classifying objects is important, as the vehicle needs to predict the future locations of surrounding objects. The systems can use such classifications to make better estimations about the future locations, knowing, for instance, that an object identified as a motorcycle is likely to be faster than one identified as a bicycle.[129] The future location of a car which could go to sixty miles per hour is likely to be very different than a bicycle likely to travel no more than twenty miles per hour.[130]

In sum, autonomous vehicles are able to navigate a complex driving environment by relying upon computer systems, which use a mix of machine learning algorithms that infer how to drive on their own, as well as relying upon some explicit computer rules. Later, the point will be made that, because machine learning algorithms operate by encoding patterns detected in data in complex computer models, it is often more difficult for people to understand how machine learning algorithms actually work. This difficulty in understanding machine learning models sometimes makes comprehending or predicting the

126 *Id.*

127 Self-driving cars are also trained virtually on computers that simulate driving scenarios as well as on the road. *See, e.g.*, Davies, *supra* note 73.

128 David Michael Stavens, *Learning to Drive: Perception for Autonomous Cars* (May 2011) (unpublished Ph.D. dissertation, Stanford University, http://www.cs.stanford.edu/people/dstavens/thesis/David_Stavens_PhD_Dissertation.pdf).

129 Mark Harris, *New Pedestrian Detector from Google Could Make Self-Driving Cars Cheaper*, IEEE SPECTRUM (May 28, 2015, 4:00 PM), http://spectrum.ieee.org/cars-that-think/transportation/self-driving/new-pedestrian-detector-from-google-could-make-selfdriving-cars-cheaper.

130 *See id.*

future behavior of systems operating under machine learning algorithms more difficult.

b. Coordination and Planning

Ultimately, all of the information from the sensors, maps, and machine learning and other algorithms are combined and orchestrated by a supervisory system whose job is to determine what action (if any) the autonomous vehicle should take next in light of the vehicle destination goals and the vehicle's surroundings.

5. Acting: Moving the Vehicle According to Plan

Finally, in the acting phase, the autonomous vehicle actually carries out the driving actions that are consistent with the supervising computer's plan. The central computer is capable of activating and controlling the major movement subsystems of the vehicle. The vehicle might accelerate, brake, steer, or stay in place, depending upon the central supervising computer system's decision. As indicated, the process of perceiving the environment, planning, and acting is a continual one that involves constant reassessment. This ensures that moment-to-moment movement decisions remain safe, legal, and desirable in the context of a rapidly changing driving environment.

II. UNPREDICTABILITY OF AUTONOMOUS VEHICLES

In February 2016, a self-driving vehicle from Google collided with a bus.[131] This was the first accident that was attributable to a fully autonomous vehicle from Google. The company's self-driving cars had been in previous accidents, but none had been caused by the vehicle's self-directed movements.[132]

---

131 Dave Lee, *Google Self-Driving Car Hits a Bus*, BBC NEWS (Feb. 29, 2016), http://www.bbc.com/news/technology-35692845.

132 Those earlier accidents were caused by human drivers of other vehicles that happened to collide with Google's self-driving car. Jennifer Elias, *Google Accepts Responsibility for First Time in Self-Driving Crash*, SILICON VALLEY BUS. J. (Feb. 29, 2016, 3:07 PM), http://www.bizjournals.com/sanjose/news/2016/02/29/google-accepts-responsibility-for-first-time.html. There was also a well-publicized fatality in 2016 involving a Tesla automobile, but it is important to emphasize that the Tesla accident did not involve full autonomy. Rather, that accident involved "L2" partial autonomy with highway lane-keeping and collision avoidance. *See* Keith Naughton, *Google's Driverless-Car Czar on Taking the Human Out of the Equation*, BLOOMBERG BUSINESSWEEK (Aug. 4, 2016), http://www.bloomberg.com/features/2016-john-krafcik-interview-issue.

In analyzing the accident, a Google spokesperson remarked,

> This is [an] . . . example of the negotiation that's a normal part of driving—we're all trying to predict each other's movements. . . . Our car had detected the approaching bus, but predicted that it would yield to us . . . . And we can imagine the bus driver assumed we were going to stay put.[133]

The inability to reliably predict behavior thus played a critical role in this accident. The human bus driver was unable to discern the future actions of the self-driving car, and the self-driving car was unable to predict the actions of the human driver.

More broadly, this accident exemplifies a larger issue of unpredictability in self-driving vehicles that may lead to conflicts between autonomous vehicles and other drivers, pedestrians, and cyclists. This Part will raise the problem of unpredictability, explore why people sometimes find computer-controlled actions difficult to anticipate, and examine some approaches to mitigating these issues in the self-driving car context.

A. *Overview*

Although predictability plays a large role in law,[134] in one particular context the literature has surprisingly little to say: predicting the movements of other people. This is because by and large (with some notable exceptions) the behavior of other people tends to be *broadly* predictable. Humans have evolved cognitive systems that allow us to reliably predict the near-term movements of those around us.

By contrast, when it comes to assessing the future actions of *machines*, people do not possess comparable intuitive abilities. This has generally not been a problem until this point, as the large automated machines moving in our physical environment—such as elevators, escalators, and factory equipment—tend to be restricted in their range

---

133 *Google Self-Driving Car Project Monthly Report*, GOOGLE (Feb. 2016), https://static.googleusercontent.com/media/www.google.com/en//selfdrivingcar/files/reports/report-0216.pdf.

134 Predictability plays an important role in law. Lawyers routinely predict the outcomes of hypothetical and actual legal cases. In legal doctrine, predictability often plays a central role as well. Also, in negligence law, defendants can be held liable for the foreseeable (i.e., predictable) consequences of their careless actions. For instance, if a driver carelessly rear ends an ordinary car that unexpectedly had radiological materials in the trunk, the defendant could be liable for the typical injuries associated with the vehicle accident (e.g., whiplash), but under the doctrine of proximate cause, not liable for consequences completely atypical and disproportionate from a minor traffic accident, such as the radiation poisoning of the surrounding neighborhood. *See, e.g.*, Benjamin C. Zipursky, *Foreseeability in Breach, Duty, and Proximate Cause*, 44 WAKE FOREST L. REV. 1247, 1249–50 (2009).

of movement and therefore broadly predictable. However, this issue will soon become more pressing as autonomous vehicles, and other self-directed machines with relatively unrestricted movement, become a common feature in our shared physical spaces.

The predictability of autonomous vehicles presents somewhat of a paradox on the surface. Autonomous vehicles are likely to be *more predictable* than human drivers in many instances, unfailingly following their instructions and the rules of driving.[135] Human drivers are less predictable in this sense because they sometimes drive while distracted, impaired, or do not follow traffic rules.[136] Similarly, from an engineering perspective, these machines may be more predictable because they are designed to react in highly predictable ways under specific conditions. However, the issue, is not predictability from a technological reliability or systems engineering standpoint, but rather from the the intuition of ordinary people who work, live, and move in the same physical proximity of moving autonomous vehicles.

In the near future, lay people—as opposed to trained specialists—will be for the first time, operating in close proximity to a variety of computer-controlled, self-directed moving systems that are physically unrestricted in their movement and have latitude to control their own movements.[137] For instance, as of the writing of this Article, self-driving vehicles are beginning to work alongside human workers in Australia in the mining industry and in Florida in the construction industry.[138] Simply by observing an autonomous vehicle, a lay person cannot easily tell what data the vehicle has gathered, what, among many sensors, the system is paying attention to, nor what behavior the internal control algorithms will instruct the system to do next. The internal states of self-

---

135 Kasey Panetta, *Why Humans Are the Problem with Autonomous Cars*, ECN, https://www.ecnmag.com/blog/2015/09/why-humans-are-problem-autonomous-cars ("The [automated] car struggles to interpret the erratic (and technically incorrect) driving habits of human drivers.") (last visited Sept. 2, 2016).

136 *See* Aviva Rutkin, *Autonomous Cars Are Learning Our Unpredictable Driving Habits*, NEW SCIENTIST (Aug. 26, 2015), https://www.newscientist.com/article/mg22730362-900-autonomous-cars-are-learning-our-unpredictable-driving-habits/?utm_source=NSNS&utm_medium=SOC&utm_campaign=twitter&cmpid=SOC%7CNSNS%7C2015-GLOBAL-twitter (describing how human drivers can be difficult to predict).

137 Research concerning robots and autonomous systems interacting with ordinary people is sometimes termed "social robotics." *See* THOMAS BOCK & THOMAS LINNER, ROBOT ORIENTED DESIGN: DESIGN AND MANAGEMENT TOOLS FOR THE DEPLOYMENT OF AUTOMATION AND ROBOTICS IN CONSTRUCTION 119 (2015).

138 *See* Paul A. Eisenstein, *Driverless Construction Zone Truck Due to Hit the Road This Year*, NBC NEWS (Aug. 25, 2015, 11:23 AM), http://www.nbcnews.com/business/autos/driverless-construction-zone-truck-due-hit-road-year-n415531 (describing autonomous vehicles used in construction in Florida); David Robson, *The Truth About Driverless Vehicles*, BBC (Oct. 13, 2014) http://www.bbc.com/future/story/20141013-convoys-of-huge-zombie-trucks (describing autonomous vehicle mining operations in Australia).

driving vehicles are neither transparent nor intuitively comprehensible to ordinary people unless specifically engineered otherwise.

This Subpart will explore how we as humans use, among other things, a theory of mind to reliably understand and predict the actions of other people, and thereby avoid a significant amount of physical harm that might otherwise occur, as we navigate an environment filled with people and machines operated by people.[139] These cognitive mechanisms are likely to be less effective, for a variety of reasons, when movement decisions are made by computers, rather than other people.

B. *Predicting the Behavior of Other People*

People work and live in close proximity to one another. To avoid injury, it is important to be aware of our movements and how they affect others around us. For instance, on a crowded bus, if a passenger were to suddenly extend her arm, this could result in injury to nearby passengers. Fortunately, we can usually rely upon others to follow social norms and be aware of their surroundings and thereby avoid such harmful interactions. More generally, we presume that other people in our environment will sense, think, and act like us in a broad sense, operating according to a few basic rules such as not wanting to injure others and acting to preserve their own safety.[140] We are generally able to coexist safely by relying upon our perceptions about how others are likely to act.

The way in which people are able to reliably predict the ordinary actions of others around them is not fully understood. It likely involves a combination of prior experience, background knowledge about the world, observation, communication, belief and knowledge of social norms, and internal analysis. However, one important part of the process is thought to be a series of cognitive mechanisms that allow people to estimate the internal mental states, and likely future actions, of others.

1. Theory of Mind

Researchers have termed the human ability to reliably assess the mental states, motivations, beliefs, and future conduct of other people as

---

139 D. M. Wolpert, K. Doya & M. Kawato, *A Unifying Computational Framework for Motor Control and Social Interaction*, 358 PHIL. TRANSACTIONS OF THE ROYAL SOC'Y B: BIOLOGICAL SCI. 593 (2003).

140 PHILIP E. TETLOCK & DAN GARDNER, SUPERFORECASTING: THE ART AND SCIENCE OF PREDICTION (2015).

a "Theory of Mind."[141] Such theory of mind mechanisms developed because humans are social creatures who evolved to live cooperatively in groups.[142] Crucial to the ability to function in a collective environment was the ability to create social relationships. Such social bonds required the ability to understand what others in the group were thinking about, feeling, and paying attention to. For this reason, researchers believe that humans evolved cognitive facilities capable of assessing the current and future mental states and physical actions of other people around them through observation, introspection, and projection.[143] Although the term "theory of mind" is sometimes used narrowly in the research literature to refer to predicting the mental states of others, this Article will use the term broadly to refer to the collective set of cognitive mechanisms that allow people to predict both the physical and mental states of others.[144] In addition to predicting movement, a theory of mind is fundamental to human-to-human communication and the development of human language.[145]

The possession of theory of mind abilities allows for implicit social ordering and harm avoidance. Based upon extrapolating from our own beliefs and motivations, and relying on instinct, we are often able to reliably predict what others around us are likely to do. For instance, as we drive, we expect and assume that others will follow social conventions and legal rules, such as staying on the correct side of the road and not veering into oncoming traffic.[146] There are a complex series of reasons for why we might collectively choose to follow such a critical driving rule like this, among them are likely: desire to minimize risk to ourselves and preserve our own safety (e.g., recognizing the severe danger of driving into oncoming traffic), a desire to avoid chaotic driving, and motivation by fears of legal sanctions and violating social norms.

More generally, as we drive on roads, we implicitly trust that others will mostly follow certain critical conventions as well. Our internal theory of mind mechanism gives us the ability to understand when

---

141 Premack & Woodruff, *supra* note 8, at 515.

142 Siegal & Varley, *supra* note 9.

143 For social creatures living in groups, it is important not just to be able to assess the mental states of others, but also their future physical states. Traveling in groups raises the risk of harmful physical collisions, and it is advantageous to be able to avoid physical collisions where possible. For this reason, we have also developed mental systems capable of understanding and predicting the future movements of those around us, and avoiding collisions. For this reason, this Article will use the term "theory of mind" broadly not just to refer to the ability to assess the mental states of others, but also the ability to assess the future physical movements of others. Beetz, Johnston & Williams, *supra* note 7.

144 Alvin I. Goldman, *Theory of Mind*, *in* THE OXFORD HANDBOOK OF PHILOSOPHY AND COGNITIVE SCIENCE 402 (Eric Margolis, Richard Samuels & Stephen P. Stich eds., 2012).

145 *Id.*

146 TETLOCK & GARDNER, *supra* note 140.

others are likely to follow norms because they are likely internally motivated by similar social, legal, and safety-preservation goals as we are. The ability to predict the behavior of others by understanding when they share similar concerns and beliefs is one aspect of the theory of mind that helps to reduce physical harm and risk of collision.

Theory of mind facilities also allow us to react to dynamic and changing circumstances in our immediate physical environment. Humans have the ability to observe the facial expressions, gestures, and movements of those around us, interpret those signals, and react accordingly. For instance, people are able to walk through dense crowds without generally colliding with one another. Part of this ability relies upon our understanding that others will follow certain unstated social rules, such as not intentionally colliding with others in their path. But another part of this remarkable ability to navigate seemingly chaotic crowds with little or no physical injury stems from our innate ability to assess the movements of other people with whom we might collide, discern where they are likely to move, and redirect our own movements.

Our internal mental machinery is capable of picking up subtle bodily cues from others about the direction of likely movement, such as whether a person is likely to move left or right.[147] For example, people tend to lean slightly in the direction that they are about to move in, and our cognitive mechanisms can detect and process these small movements.[148] We can then move in the opposite direction if it looks like we are on a collision course with an oncoming person. We have the ability to observe and react accordingly, adjusting our own movements.[149] Such assessments and predictions occur nearly instantaneously and generally below our consciousness. This is necessary to allow timely reactions.

As discussed, such predictive analysis is critical to safety in the driving context. Let us return to the earlier example of the pedestrian about to enter a crosswalk with an approaching human-driven vehicle. It is crucial to the pedestrian's safety that she be able to predict whether the driver of the vehicle is likely to stop. The pedestrian will use her internal theory of mind cognitive capacities to make a series of rapid assessments: What is the driver paying attention to? Is he looking ahead at the road, or down at his cell phone? What are the driver's capabilities? Does his vehicle appear to be capable of stopping through normal

147 Peter Collett & Peter Marsh, *Patterns of Public Behavior: Collision Avoidance on a Pedestrian Crossing*, *in* NONVERBAL COMMUNICATION, INTERACTION, AND GESTURE: SELECTIONS FROM SEMIOTICA 199, 215 (Adam Kendon, Thomas A. Sebeok & Jean Umiker-Sebeok eds., 1981) (exploring how people pick up on non-verbal cues that people give out that indicate their direction of movement).

148 Kendon, *supra* note 3, at 18.

149 Moussaïd et al., *supra* note 11.

braking, or is it careening abnormally out of control? Does the driver appear to be intending to stop, or does he appear to be in a hurry and ready to dash through the intersection?

Using theory of mind mechanisms, the pedestrian will extrapolate from her own experience and capabilities and project them onto the driver in order to make a prediction about his likely behavior.[150] If the pedestrian sees the driver looking down at his cell phone, she will intuit that the driver's attention is directed away from the pedestrian. The pedestrian can then predict that the driver is unlikely to stop and avoid stepping in front of the vehicle. In general, predicting the behavior of others is important to avoiding certain types of harm, and people routinely rely upon theory of mind mechanisms to make such predictions in the driving context.

2. Communication

Communication is also crucial to making better predictions about driver and pedestrian behavior. As people drive, they verbally and non-verbally communicate with those around them to signal their intentions.[151] As discussed, drivers and pedestrians sometimes make eye-contact or wave, communicating to one another that they have been perceived. Knowing whether or not they have been detected by drivers is a crucial point of safety for pedestrians, cyclists, and other vulnerable populations.

However, as the Google accident illustrated, communication is also crucial in the driver-to-driver context. Driving is a constant and dynamic negotiation with other drivers on the road. A driver who is trying to merge may wave to get the attention of another driver to move in front. Drivers use turn signals to indicate when they are changing lanes or turning. Similarly, communication is also key in the driver to passenger context. A driver who is too tired to drive may indicate as such, or the passengers may observe as much. In general, communication between drivers and those around them—other drivers, passengers, and pedestrians—is a crucial component of safe driving, and theory of mind mechanisms allow us to better predict human behavior through verbal and non-verbal communication.[152]

---

150 *See* Goldman, *supra* note 144.

151 *See, e.g.*, Rutkin, *supra* note 136.

152 *Id.*

C. *Technological Opacity and Unpredictability*

In a sense, having a theory of mind allows us to peer into the minds of other people from afar.[153] We can distill the inner working and mental states of those around us without having direct physical access to their brains. Possessing such an internal mental model of the way people operate allows us to make reliable predictions about the beliefs, mental states, actions, and intentions of others. By contrast, the activities of technological systems such as autonomous vehicles are not susceptible to this type of modeling through introspection. Our theory of mind cognitive mechanisms evolved to predict the near-term actions of people and not the behavior of machines. In other words, humans have no instinctual basis for understanding what an algorithmically controlled technological system such as an autonomous vehicle is going to do next. More generally, people do not have innate mental models that allow them to externally discern the internal states of technological systems, or to communicate with these systems to convey crucial information.

Autonomous vehicles make decisions about where to move based upon data gleaned through their sensors and analysis by computer algorithms. We can refer to the combination of data and computer analysis that an autonomous vehicle is relying upon to make decisions as its "internal state." Like most technological systems, such as computers or smartphones, the internal state of such a system—its data and software—are stored in machine-friendly, but not human-comprehendible form, as electronic data.

The problem is that the internal states of such technological systems are simply not transparent to ordinary people. Such systems have to be expressly designed to convey their intentions to people meaningfully, and as of today, autonomous vehicles are not fully designed to do so.[154] This lack of internal transparency and predictability is not unique to autonomous vehicles, but rather applies to some degree, to most electronic technologies. However, this issue becomes more important in the context of autonomous vehicles because they are physically large objects with free-ranging, self-directed movements capable of seriously injuring people.

---

153 Goldman, *supra* note 144.

154 *But see* Ben Popper, *A New Patent Reveals How Google's Self Driving Cars Could Talk to Pedestrians*, VERGE (Nov. 27, 2015, 8:00 PM), http://www.theverge.com/2015/11/27/9808658/google-driverless-car-patent.

1. Technological Opacity

Let us call a computer-based system "technologically opaque" if it is difficult for an ordinary person to understand why that technological system takes a particular action that it does.[155] A good example of technological opacity comes from aviation. For many years commercial jets have had sophisticated autopilot systems that automate many of the tasks involved in flying.[156] These systems are extremely complex and lend a significant amount of automated assistance to steering, navigation, landing, and other core flying activities. Such automated autopilot systems are believed to have substantially improved overall aviation safety.[157] Auto-piloted airplanes today are able to routinely land safely in dangerous conditions—such as dense fog—that were previously difficult for human pilots.[158]

However, these automated systems are sometimes technologically opaque to the pilots who use them. As experts have observed, it is not uncommon for pilots in the cockpit to be surprised or confused by an automated activity undertaken by an autopilot system. This has been captured in a common industry catch-phrase, "What [is the system] doing now?" in reaction to an autopilot's unexpected activity such as suddenly changing the airplane's altitude.[159] An autopilot can, thus, undertake automated actions that, even if safe and appropriate for the conditions, may not be readily understandable or intuitive to the pilots.[160]

We can, thus, characterize a system as technologically opaque if it engages in automated actions whose basis is difficult for human users to

---

155 The phrase "technological opacity" is our own terminology. For discussions of opacity in other contexts, see Frank Pasquale & Danielle Keats Citron, *Promoting Innovation While Preventing Discrimination: Policy Goals for the Scored Society*, 89 WASH. L. REV. 1413, 1422 (2014); Omer Tene & Jules Polonetsky, *To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising*, 13 MINN. J. L. SCI. & TECH. 281, 296 (2012).

156 Simon Wood, *Flight Crew Reliance on Automation*, CIV. AVIATION AUTHORITY (2004), https://publicapps.caa.co.uk/docs/33/2004_10.PDF.

157 FED. AVIATION ADMIN., ADVANCED AVIONICS HANDBOOK ch. 4 (2009).

158 *Id.*

159 *See* Lance Sherry et al., *What's It Doing Now?: Taking the Covers Off Autopilot Behavior*, RESEARCHGATE (2001), https://www.researchgate.net/profile/Lance_Sherry/publication/228761489_What's_it_doing_now_Taking_the_covers_off_autopilot_behavior/links/00b7d5294c1cc0b936000000.pdf; Katie Mingle, *Children of the Magenta (Automation Paradox, Pt. 1)*, 99% INVISIBLE (June 23, 2015), http://99percentinvisible.org/episode/children-of-the-magenta-automation-paradox-pt-1 (discussing the phenomenon of automation and the inability for ordinary people to understand certain automated decisions); *see also* 99% INVISIBLE, *Air France Flight 447 and the Safety Paradox of Automated Cockpits*, SLATE (June 25, 2015, 8:51 AM), http://www.slate.com/blogs/the_eye/2015/06/25/air_france_flight_447_and_the_safety_paradox_of_airline_automation_on_99.html.

160 99% INVISIBLE, *supra* note 159.

understand. It is important to distinguish such technological opacity from a system malfunction that results in erroneous behavior, such as a software bug or computer crash. Technological opacity applies when a technological system is functioning exactly as intended under the given conditions, but it is simply not transparent to a person why the system took the automated actions that it did.

More broadly, "technological opacity" applies any time a technological system engages in behaviors that, while appropriate, may be hard to understand or predict, from the perspective of human users.[161] This is a common phenomenon and generally related to the underlying complexity of modern technological systems. Part of the reason why an airplane autopilot system may be technologically opaque may be due to the complicated nature of the endeavor. Flying is a difficult task, so a system capable of automating flying is necessarily going to be complex. The physical parts, subcomponents, data, and software interact in intricate ways to produce the desired automation. In the case of such a complicated system, it may be difficult to convey what that system is doing and why, in a manner that is meaningful to people.

Similarly, most modern technologies, from smartphones to autonomous vehicles, are composed of an intricate mix of electronic components, sensors, data, and computer software. These various components interoperate in complicated ways that are well understood to the engineers who designed them. However, often the design goal of engineering is to hide this underlying complexity from the end-user to make it useable, and to present only the most useful information. For instance, people do not need to understand the underlying electronics, data, and software in a smartphone in order to use the device. Such phones are operated by the user through a simple graphical user interface. However, the consequence of masking the underlying complexity sometimes results in technological opacity: an ordinary end user may not intuitively understand why a technological device acted the way it did, nor what it is going to do next.

Most technological systems created for end users are designed to communicate key information through *interfaces.* Interfaces are means of communicating system information to a human user, and include screens, visualizations, graphical user interfaces, dials, gauges, lights, sounds, reports, and other methods. Engineers design systems with interfaces to reveal information regarding the internal states of machines in ways that are relevant to users.

---

161 *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1717 (2010) (describing how difficult it is for researchers to understand the level of user anonymity in complex data).

By default, most technological systems do not reveal their internal working states externally. Rather, such systems must be deliberately designed to communicate relevant information externally. Engineers make explicit decisions about what information to convey outwardly to users and what not to reveal. Thus, most information that has not been explicitly designed to be communicated will tend to be inaccessible, non-comprehendible, or technologically opaque to ordinary users.

In this sense, the term "technological opacity" is meant to be more granular than the concept of a "black box."[162] The term "black box" usually suggests a technological system which is unknowable from the outside, or whose internal details are deliberately hidden, whereas "technological opacity" suggests a *spectrum* along which systems can be designed to be incrementally more outwardly transparent about their decision-making processes or planned activities. In sum, to the extent that crucial functionality information is not explicitly communicated externally, the automated actions of technological systems may be difficult to understand and predict, and they may therefore be technologically opaque.

2. Current Autonomous Vehicles Are Technologically Opaque

Autonomous vehicles, as they are currently designed, tend to be technologically opaque in certain respects, and therefore less predictable to ordinary people. Such vehicles combine an elaborate mix of sensors, electronic components, mechanical components, computer systems, symbolic models, and controlling software in order to drive. Like the autopilot of a commercial airplane, the operation of such a complex electro-mechanical system is not intuitive to lay people. But beyond this complexity are a few nuances worth pointing out relating to the functionality of a typical self-driving vehicle.

Self-driving cars rely upon their sensor data to avoid colliding with pedestrians or others. However, it is not always transparent to external observers what, among many nearby objects, an autonomous vehicle has detected with its sensors. Sensor information is directed into the vehicle's central planning system to determine where it is safe, legal, and desirable to move next, while avoiding obstacles. (In some cases, this information is also displayed internally to the autonomous vehicle's passengers.) However, in general, this information about the vehicle's internal state is not communicated externally or meaningfully to those

162 For an excellent discussion of the "black box" concept, see FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

surrounding people—other drivers or pedestrians—who may need it the most.

For example, it is often important to cyclists and other vulnerable entities to firmly determine whether or not they have been detected by nearby vehicles. Due to the technological opacity of current self-driving car designs, it may be difficult for a cyclist to reliably determine whether or not the vehicle's sensors have detected her. More broadly, engineers and designers have not given sufficient thought to usefully communicating to nearby vulnerable individuals (pedestrians, cyclists, and other drivers) what the vehicle has actually detected with its sensors.

Another related issue concerns the sensor information the system is paying attention to.[163] Consider our bicyclist example once more. Imagine that one of the vehicle's sensors has indeed detected the cyclist. One might assume that this is the end of the inquiry, that the vehicle will seek to avoid hitting her, the way a human driver would. However, even if the cyclist is perceived by a sensor, there is no guarantee that the vehicle's central computer will *prioritize* this particular piece of information and act upon it. As they drive, autonomous vehicles take huge volumes of information from multiple sensors. Any time a technological system receives large amounts of data, it must have a method for sorting through such information to determine what to pay attention to and prioritize and what to ignore. The distinction between an autonomous vehicle sensing a nearby cyclist, and paying attention and prioritizing this signal, is a subtle but important one and relevant to safety. Imagine, for instance, that in addition to detecting the cyclist, the vehicle has also detected a pedestrian in front of the vehicle, and has prioritized avoiding the pedestrian over the cyclist, or even the passengers onboard.

Thus, even if the vehicle communicates to the cyclist that she has been perceived—this may not be enough. The cyclist must be confident that not only will her presence be perceived by the vehicle, but that the vehicle will act in a way consistent with avoiding a collision with the cyclist. Such an "attention architecture" of the vehicle's computer system—the types of information the system thinks is important, is paying attention to, and is likely to react to—is not inherently obvious to outsiders, and needs to be communicated.[164]

163 Novianto, Johnston & Williams, *supra* note 26.
164 *Id.*

3. Machine Learning and Comprehensibility

Self-driving cars can also be technologically opaque due to their software. As described earlier, self-driving vehicles use, to some extent, a programming technique known as machine learning. Machine learning can present some specific challenges in terms of predictability compared to more traditional programming techniques. As discussed, machine learning differs in a significant way from the instruction-based programming approach that underlies most computer software.[165] In the traditional "explicit" programming process, programmers create software through a series of computer instructions. A computer then systematically follows those instructions.[166] By contrast, machine learning systems are developed by a different method in which the software effectively programs itself by analyzing large amounts of data to look for useful patterns. These patterns are then encoded as models—formulas or other complex data structures—that are well-suited for computers (but not people) to follow. The computer then uses the complex patterns that it has detected to make automated decisions involving new data, and this process allows it to engage in very sophisticated automated tasks, such as driving.

A major difference between machine learning and "traditional" programming is thus the explicitness of the rules upon which the computer makes its decisions. In a traditional computer program, because the computer is following a clear list of instructions written by a person, which can be inspected and understood, it is relatively easy for a programmer to understand why a computer made a particular decision that it did. By contrast, in machine learning, the computer is often following a highly abstract pattern gleaned from analyzing huge troves of data.

Because of the complexity and abstractness of machine learning models, even the programmers who created them are not always able to understand how and why they perform the way that they do. Computer scientists sometimes refer to this as the "comprehensibility principle."[167] It is common to have a well-functioning machine learning system that makes appropriate decisions about a task such as driving, but whose inner logic is not readily comprehensible. In other words, even if one has a machine learning model that works well in practice, a programmer

165 Surden, *supra* note 121.

166 *Id.*

167 Ryszard S. Michalski & Yves Kodratoff, *Research in Machine Learning: Recent Progress, Classification of Methods, and Future Decisions*, *in* MACHINE LEARNING: AN ARTIFICIAL INTELLIGENCE APPROACH, VOLUME III 3, 6 (Yves Kodratoff & Ryszard S. Michalski eds., 2014).

may still not be able to understand the underlying reasons that allowed it to produce good results.

Machine learning algorithms are thus often evaluated on their functionality—how well they perform at a particular task such as driving—rather than on their understandability. However, because it is not always easy to know why a machine learning algorithm made a particular decision that it did, and because the inner logic of such systems is not necessarily revealed through inspection, this lack of intelligibility can impact the predictability of such systems, from the perspective of lay observers.

A second, related issue is that machine learning programs are designed to "learn" over time and change how they act as they encounter new data. It is certainly possible to train a machine learning model and then "freeze" it so it does not change over time as it encounters new data. But some types of machine learning algorithms do change their own programming over time as new data becomes available. Such an ability to change and learn can be beneficial, as it can lead to better driving behavior.[168] However, the ability for software to change its own programming over time as it encounters new data might add to the difficulty in predicting self-driving car behavior, as the vehicle will essentially have different (and perhaps differently behaving) software over time.

In sum, the implicit, dynamic, and abstract nature of machine learning software—as contrasted against the explicit and linear nature of traditional programming by explicit computer instruction—may make it more difficult to communicate, in a meaningful way to others nearby what action an autonomous vehicle has decided to take and why.

III. IMPLICATIONS FOR LAW

The emergence of fully autonomous vehicles presents challenging and novel issues for the legal system. As described in the last Part of this Article, decisions made by computers tend to be less intuitively predictable from a human cognitive level than comparable decisions made by people. People can predict the activities of one another by relying upon internal models of human behavior and by signaling their intentions through communication. This Part suggests that the activities of autonomous vehicles (and other computer controlled autonomous systems such as robots) can be made more predictable through deliberate technological design decisions. In general, the topic of making autonomous moving systems more outwardly predictable has

168 Fehrenbacher, *supra* note 66.

simply not received sufficient attention.[169] This Part also highlights particular scenarios that are worthy of attention, and explores to what extent the legal system might (or might not) be involved in encouraging increased technological predictability.

A. *Unarticulated Assumptions of Predictability in Law*

Our current legal structure contains unarticulated presumptions of movement predictability. Within tort law, the doctrine of comparative fault penalizes a plaintiff for not preventing or reducing the injuries from an accident that were clearly avoidable.[170] For example, if a pedestrian carelessly runs in front of an approaching human-driven vehicle and is injured, the pedestrian's damage award may be reduced or eliminated. This is because we believe that the pedestrian could have avoided or mitigated the accident had she not been careless. Although it is not often stated, embedded in the notion of an avoidable injury is the assumption that the plaintiff could readily predict the behavior of the injuring party. We expect a person to know that dashing in front of a moving car may not give the driver enough time to stop.

But contrast a similar scenario in the context of a computer-directed autonomous vehicle. Can a pedestrian reasonably be expected to predict the activities of an approaching autonomous vehicle in the way she might reliably predict the actions of a human-driven vehicle? Perhaps a self-driving vehicle could be interpreted as sending indications that it was going to stop. Is it reasonable to say that the pedestrian took a contributing risk in darting in front of a self-driving car when she may not have been aware how an autonomous vehicle was going to react, or the underlying technological capabilities of the vehicle? Notions of predictability embedded in tort law may be challenged when activities are made by self-directed moving autonomous systems whose computer controlled actions may be difficult for ordinary people to anticipate. Is a pedestrian careless for failing to avoid an injury from a self-driving car, whose movements may not be instinctively predictable, but that she had plenty of time to avoid?

This is one example of a broader thread through much of law that seems to be based upon a presumption that the movements in our physical environment will be made by other *people* with similar goals, desires, and perceptions, and will therefore be broadly predictable.

---

169 Most of the current research self-driving car communication is focused on communicating information to those inside the car—the driver and passengers—and not to pedestrians, cyclists, and other external individuals. *See, e.g.*, NHTSA, *supra* note 48, at 7 (discussing communication in the context of human drivers in self-driving cars).

170 *See, e.g.*, Am. Motorcycle Ass'n v. Superior Court, 20 Cal. 3d 578 (Cal. 1978).

However, as discussed, such a presumption may be less accurate when movement decisions are instead made by computers. The legal system has not had to grapple with such a question because until the development of self-driving cars, computer-directed, free-ranging movement of large machines in public had simply not been an issue.

As I have written elsewhere, this is part of a reoccurring pattern involving emerging technology.[171] We can often think of societal activities as being *implicitly* regulated by the technological limitations of the past, only to have a new technology emerge and render that implicit regulation suddenly ineffective.[172] For instance, in the era of paper documents, privacy was implicitly protected by the sheer difficulty of accessing distant private information stored in paper documents and aggregating that data in useful ways.[173] Digital electronic documents and remote internet access reduced the costs of accessing, searching, analyzing, and linking previously disparate private data, and thereby removed the implicit privacy protection of physical separation provided by paper technology.[174]

Under one interpretation, the pre-digital legal privacy framework *actually relied upon* the difficulty of accessing and analyzing data on paper in order to effectively safeguard privacy. In other words, the technological cost and difficulty of accessing and analyzing paper-based data was not simply a byproduct of the primitive technology of the past, but was actually playing a crucial, but implicit, role in protecting privacy. There is an analogous technological dynamic with the emergence of self-driving cars. Our existing legal structure may implicitly depend upon our current (pre-autonomous vehicle) technological state in which we can broadly predict the movement of others. The switch from human-driven to computer-driven vehicles may undermine the embedded safety regulation of our current technological era.

When such implicit protections erode due to technological change, one approach is to actively replicate those lost protections using some mode of regulation, such as law, technology, norms, or economics.[175] For instance, in response to loss of implicit privacy protections with the

171 *See* Harry Surden, *Technological Cost as Law in Intellectual Property*, 27 HARV. J.L. & TECH. 135 (2013) [hereinafter Surden, *Technological Cost as Law in Intellectual Property*]; Harry Surden, *Structural Rights in Privacy*, 60 SMU L. REV. 1605 (2007) [hereinafter Surden, *Structural Rights in Privacy*].

172 Surden, *Technological Cost as Law in Intellectual Property*, *supra* note 171; Surden, *Structural Rights in Privacy*, *supra* note 171.

173 *See generally* Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661 (1998) (describing that these are different, sometimes complementary, modes of regulating social behavior).

174 *Id.*

175 *Id.*

shift to digital court documents, governments responded by consciously applying technological measures, such as encryption, to protect privacy in ways that were comparable to the implicit protection of the pre-digital paper era.[176] Similarly, these next Sections will discuss such an approach to actively re-architecting movement predictability in the context of self-driving cars, using both technological and legal mechanisms.

B. *Making Autonomous Vehicles More Predictable*

Like other complex technological systems, autonomous vehicles can be designed to interact more intuitively with people. Within engineering and design, there are a number of disciplines focused upon making technology more understandable and useable to ordinary users. Most directly, the research areas of human factors engineering (HFE) and human computer interaction (HCI) have specific frameworks and methodologies aimed at producing complex systems that are designed be more useable, intuitive, and informative.[177]

A good recent example of such deliberate design comes from the area of smartphones. Despite their superficial outward simplicity, modern smartphones are actually extremely complex devices underneath, composed of hundreds of different sensors, software systems and electronic components interacting with one another. Yet modern smartphone operating systems—such as Apple's iOS—have been designed using principles of HCI to make interaction relatively simple and intuitive. These systems are designed to communicate only the most important information to the user through graphical user interfaces, and hide the underlying technological complexity from the user to make using the smartphone a more intuitive and predictable experience.

Related to human factors engineering and HCI is the emerging field known as "social robotics."[178] This discipline recognizes that humans, robots, and other autonomous systems are beginning to share the same physical spaces (e.g., today factories and research laboratories) and focuses upon unique issues arising out of such increasing

---

176 *Id.*

177 For a definition of HCI, see HEWETT ET AL., ACM SIGCHI CURRICULA FOR HUMAN-COMPUTER INTERACTION ch. 2.1 (1996), http://old.sigchi.org/cdg/cdg2.html ("Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.").

178 Mary-Anne Williams, *Robot Social Intelligence*, *in* SOCIAL ROBOTICS: 4TH INTERNATIONAL CONFERENCE, ICSR 2012, CHENGDU, CHINA, OCTOBER 2014 PROCEEDINGS 45–55 (Shuzhi Sam Ge et al. eds., 2012).

interactivity between humans and moving autonomous systems. There also exists an analogous discipline to HCI in the field of robotics, Human Robotics Interaction (HRI),[179] but that field is far less mature and generally lacks design principles and frameworks.[180]

Collectively, these fields recognize that the relationship between technological systems and the people who interact with them can be made better (or worse) through deliberate decisions about how they are designed. Most of these fields consider principles of how humans naturally think, communicate, operate, and process information in order to make complex technologies more intuitive, trustworthy, expressive, and useable.

This Section will not attempt to specify particular technological design solutions to make autonomous vehicles more predictable to the pedestrians, drivers, and others that share their physical space. Rather, it will identify a series of representative scenarios and some general principles—such as increased communication—that merit attention and which are illustrative of the underlying unpredictability issue in self-driving cars. Most of these scenarios involve high conflict contexts between autonomous vehicles and vulnerable populations—such as pedestrians, bicyclists, other drivers, or passengers—that should be of particular concern to policymakers.

1. Communicating to People that They Have Been Detected

Autonomous vehicles must be designed to communicate to nearby people that they have been detected. As described earlier, autonomous vehicles use a variety of sensors such as radar, lidar, sonar, and video cameras to detect and avoid obstacles such as pedestrians, cars, and cyclists. However, it is crucial that ordinary people know if and when nearby autonomous vehicles have detected their presence. This issue has been illustrated multiple times throughout this Article through the various examples involving pedestrians at crosswalks.

Many examples have expressed uncertain deadlock scenarios between an approaching autonomous vehicle and other people. In most of these scenarios, the concern has been that, under current designs, cyclists, pedestrians, or drivers faced with an approaching autonomous vehicle cannot be completely confident that they have been detected in the way they might with a human driver. To be clear, this is not a technological capability issue—it is extremely likely in most cases that

179 Hugo Romat et al., *Natural Human-Robot Interaction Using Social Cues*, 11TH ACM/IEEE INT'L CONF. ON HUMAN-ROBOT INTERACTION 503 (2016).

180 Robin R. Murphy et al., *Human–Robot Interaction*, 17 IEEE ROBOTICS & AUTOMATION MAG. 85 (2010).

an autonomous vehicle will *actually* detect a pedestrian—as vehicles have multiple redundant sensors that allow for such detection with a high probability. Rather, this represents a technological design and communication issue. A self-driving car may have in fact detected a pedestrian, but may simply not be designed in such a way so as to effectively communicate that information externally.

Human drivers encounter these same scenarios. To navigate these conflicts, they use verbal and non-verbal communications to indicate to pedestrians or other vulnerable populations that they have been perceived. Through an iterative set of explicit waves or subtle eye contacts, pedestrians and others nearby can gain more confidence that they have been perceived. Contemporary self-driving vehicles, by contrast, tend not to have comparable external communicative abilities. Autonomous vehicles need to be designed in such a way that they are able to meaningfully communicate to those around them that they have been perceived by the vehicle's sensors, and that the system's attention architecture is paying attention to them.

2. Communicating Intentions to Surrounding People

It is not enough that autonomous vehicles communicate to those around them that they have been detected. An autonomous vehicle must also clearly communicate what it is going to do next. For instance, a pedestrian at a crosswalk might be unsure if a self-driving vehicle is intending to stop and wait for the pedestrian to cross or drive through the crosswalk.

It is also important that self-driving cars be designed to communicate their intentions to other drivers. The Google self-driving car accident described earlier illustrates this issue of communicating with human drivers.[181] As Google noted, driving can be thought of as a negotiation, where drivers are constantly communicating with those nearby, to navigate safe passage.[182] In a sense, the Google self-driving car accident was a failure of communication. The Google car did not have a good means of communicating to the human bus driver what it was about to do, and the human bus driver did not have a good method of communicating to the self-driving car what it intended to do. Because of the absence of communication methods, both the human and autonomous drivers were left to use assumptions about the other's behavior, which proved faulty.

---

181 *See supra* Part II.

182 Elias, *supra* note 132.

To a limited extent, autonomous vehicles already do communicate their intentions. Just like human drivers, they employ their automobile turn signals when switching lanes or turning and activate their rear brake lights when slowing or stopping. All of these are important tools that communicate intentions that parallel the signals of human drivers. However, because the internal computer-controlled decisions of autonomous vehicles are not transparent or intuitive to people, but rather are controlled by software, data and algorithms, it may be important to provide a level of communication about the near term movements of the vehicle to those around them beyond these simple indications.

"Vehicle to Vehicle" (V2V) or "Connected Vehicle" technologies may improve this communication issue. "Vehicle to Vehicle Communication" is the name for a suite of technologies that allow cars to communicate their location, speed, and movements to other nearby cars.[183] Such communication can improve overall driving safety, as cars can wirelessly broadcast to other cars around them exactly what they are doing, allowing other drivers to make more informed driving decisions.[184] While uncommon as the writing of this Article, V2V communication is likely to become more common in ordinary cars by 2020 due to federal mandates.[185] Ultimately, these technologies can likely be used by self-driving cars to communicate their near term-autonomous movements.[186] However, V2V communication is likely to only be a comparatively small part of the solution.

Relatedly, it is not always clear what information a self-driving vehicle is paying attention to. Recall that autonomous vehicles are inundated with a flood of information from multiple sensors, detecting not only obstacles such as other vehicles, but also traffic features such as curbs, lane markings, signals, etc. The vehicle's computer system must necessarily prioritize some of this detected information, and pay less attention to other information that it has detected, given its limited computing resources. A pedestrian likely not only wants an assurance that it has been detected, but that the vehicle's computers are paying attention to and have prioritized her physical presence (as opposed to other objects detected) and is planning to avoid the pedestrian (as opposed to taking other actions).

---

183 Will Knight, *Car-to-Car Communication*, MIT TECH. REV. (2015), https://www.technologyreview.com/s/534981/car-to-car-communication.

184 *What Are Connected Vehicles and Why Do We Need Them?*, U.S. DEP'T OF TRANSP., http://www.its.dot.gov/cv_basics/cv_basics_what.htm (last visited Sept. 21, 2016).

185 Kirsten Korosec, *Obama Administration to Fast-Track "Talking" Car Mandate*, FORTUNE (May 14, 2015, 5:45 AM), http://fortune.com/2015/05/14/v2v-communication-cars.

186 *See* NHTSA, *supra* note 48, at 3–4 (discussing vehicle to vehicle technology).

Autonomous vehicles may also unintentionally send misleading signals through their computer-controlled actions. Recall the earlier example in which an autonomous vehicle had a safety rule that it automatically slowed in speed at a crosswalk, whether or not it actually detected any nearby pedestrians. Such a slowing action might be misinterpreted by a nearby person as a deliberate signal that the vehicle is intending to stop—much the way a human driver might non-verbally signal such an intention—when in fact the vehicle may not have been intending to communicate anything at all.

Finally, consider a different type of communicative conflict: what to do when an autonomous vehicle sends messages that conflict with those from the human passengers in their car? For example, imagine that an autonomous vehicle, with a human passenger riding inside, approaches a crosswalk with a pedestrian walking nearby. The autonomous vehicle announces to the pedestrian that it is not planning to stop because it considers the pedestrian to be too far away. However, the vehicle's passenger, out of politeness, waves the pedestrian across. This presents a conflict, because the passenger is not in fact controlling the vehicle's activities—the computer is, and the computer is not planning on stopping. In such a context, the pedestrian may rely upon this human assurance rather than the automated vehicular communication. Such conflicts between the messages sent by human passengers (who do not actually control the vehicle) and autonomous vehicles themselves (which do) are worth further consideration.

3. Communicating Capabilities of Autonomous Vehicles

One important issue will likely arise particularly in the early years of fully autonomous vehicles: ordinary people may not realistically understand what autonomous vehicles are capable of doing. It is, thus, important to educate the public about the actual capabilities of autonomous vehicles generally and also for vehicles to be able to communicate their specific capabilities to the people around them.

People are likely to underestimate some of the technological abilities of self-driving vehicles once they arrive. Especially in the early transitional years, people may be uncertain as to whether an approaching autonomous vehicle will avoid them, even when the vehicle has quite capably detected them. A good recent example comes from a cyclist driving alongside an experiment self-driving car.[187] Perhaps unbeknownst to the cyclist, the vehicle had been programmed to detect bicyclists and even read cyclist hand signals, such as a wave signaling an

187 Crowe, *supra* note 22.

intention to merge in front of the autonomous vehicle. However, the uncertainty over the capabilities and intentions of the autonomous vehicle lead to a standoff in which the cyclist hesitated, uncertain as to what the autonomous vehicle was going to do.[188] This type of uncertainty about the actual detection and obstacle avoidance capabilities of autonomous vehicles may lead to deadlock and safety problems. More generally, due to safety concerns, people may act unduly cautious around self-driving cars as compared to comparably situated human drivers, leading to inefficiencies.

Similarly problematic, people are also likely to overestimate the capabilities of self-driving cars. As with any new, complex, or unfamiliar technology, people may have assumptions or beliefs about what autonomous vehicles can and cannot do. However, these assumptions may not align with the actual abilities of the vehicle. Because autonomous vehicles are likely to exceed human drivers in certain areas, people unfamiliar with the technology may assume that autonomous vehicles have superior capabilities in other areas that they do not in fact have. One could imagine, for instance, a cyclist at a crosswalk deciding to dart at high speed in front of an approaching autonomous vehicle, assuming that the vehicle's advanced sensor and braking capability will allow it to stop in contexts that exceed human driver capabilities.

While this might seem like an unreasonable action on the part of the cyclist, recall that people trust the reliability and reactivity of technology in similar ways today. Many people will not hesitate to insert their arm between the closing door of an elevator to prevent it from leaving.[189] On the surface, this seems like a terribly risky action. Inserting one's limb between two crushing pieces of metal driven by a motor—when viewed in the abstract—sounds both unreasonable and excessively risky. But the mechanisms for detecting such objects have become so reliable, and the public trust in this technology so strong, that many expect them to operate with near perfect accuracy.[190] People, thus, have formed a belief about the reliability and the capabilities of a technology formed over years of interactivity. People may bring similar assumptions about the extent and capabilities of autonomous vehicles, from their interactions and experiences.[191] Not all of these beliefs will

188 *Id.*

189 LEE EDWARD GRAY, FROM ASCENDING ROOMS TO EXPRESS ELEVATORS: A HISTORY OF THE PASSENGER ELEVATOR IN THE 19TH CENTURY (2002).

190 Steve Henn, *Remembering When Driverless Elevators Drew Skepticism*, NPR (July 31, 2015, 5:08 AM), http://www.npr.org/2015/07/31/427990392/remembering-when-driverless-elevators-drew-skepticism.

191 *See* KALRA, ANDERSON & WACHS, *supra* note 53, at 21 ("Suppose that most cars brake automatically when they sense a pedestrian in their path. As more cars with this feature come to be on the road, pedestrians may expect that cars will stop, in the same way that people stick their limbs in elevator doors confident that the door will automatically reopen. The general

necessarily comport with the actual capabilities of particular autonomous vehicles.

An example of a person misunderstanding the actual capabilities of a partially autonomous vehicle illustrates the point. A human driver assumed that a partially autonomous vehicle had the type of automatic obstacle detection, avoidance, and braking capabilities described earlier.[192] Believing this to be the case, the driver intentionally navigated the vehicle towards a crowd of people, assuming that the vehicle's autonomous systems would automatically detect the people and stop in time. In fact, the driver of the vehicle was mistaken, and the vehicle was not in fact equipped with such an automated obstacle detection and braking system, leading to a collision.[193]

As the technology emerges, ordinary people are not going to be necessarily familiar with the actual capabilities of any particular autonomous vehicle. Complicating efforts, self-driving vehicles from different manufacturers are likely to have slightly different capabilities from one another. Finally, there is the possibility that consumers may modify their self-driving vehicles from the model initially delivered by a manufacturer, increasing uncertainty about capabilities further.[194]

However, it is possible to reduce accidents of the type just described through improved education as to the actual capabilities of various vehicles.[195] Similarly, it will be important to standardize vehicles so that they have a minimum set of abilities upon which people can rely, and to consider ways to communicate a vehicle's actual capabilities if they diverge from the norm.

Several common themes emerge from the preceding discussion: the importance of communication, technological design to facilitate communication, and educating the public about the actual capabilities of autonomous vehicles as they emerge. It is important that researchers think not only about how to make autonomous vehicles drive safely and accurately—as they are primarily doing now—but also about ways of communicating the following to a lay public: what autonomous vehicles have detected and what they are paying attention to; what the vehicles are about to do next, and the underlying technological capabilities in a realistic way of any given autonomous vehicle. Communicating the actual capabilities of an autonomous vehicle, and the scope and extent,

---

level of pedestrian care may decline as people become accustomed to this common safety feature.").

192 Kresge, *supra* note 70.

193 *Id.*

194 Calo, *supra* note 25.

195 Another issue is that the autonomous technologies may fail. It is important to be able to communicate when this is the case to the passenger or drivers. *See* KALRA, ANDERSON & WACHS, *supra* note 53, at 9 ("One challenge is to ensure that the driver understands when the system works properly and when it could fail or has failed.").

is both an educational problem and a design problem that is likely to prove challenging. However, once these issues receive proper attention, it is the belief that many of these issues can be overcome through deliberate technological design decisions in augmenting the communicative capabilities of autonomous vehicles.

4. Robots and Other Moving Autonomous Systems

Autonomous vehicles are sometimes referred to as "robotic vehicles" and many of the researchers developing self-driving cars have robotics backgrounds.[196] This is because many of the features of an autonomous vehicle allow one to reasonably characterize it as a type of robot. A good working definition of a robot is a computer-controlled machine that *moves* through the environment or produces physical action in the world and which has some degree of freedom about where to move.[197]

It is, thus, the ability to *move* through or influence the physical world—whether through wheels or by rotating a robotic arm—that is what generally distinguishes a robot from other computer-controlled automated technologies. By contrast, the vast numbers of automated computer systems in the world retrieve, analyze, or communicate intangible data in an automated way, but are not considered *robots* because they do not cause action in the physical world. For instance, the results returned from a Google search are automated in the sense that they are generated entirely by computer (and not a human), but this automation concerns the analysis and retrieval of data and not the production of physical movement. Robots, on the other hand, produce physical action, such as an autonomous vehicle driving on the streets, a robotic assembly arm in a factory lifting automobile parts, or a surgical robot making an incision during surgery. And like self-driving vehicles, the degree of autonomy of movement in a robot can range from fully autonomous, to semi-autonomous, to non-autonomous, depending upon the extent to which the robot controls its own activity.

This linkage to robotics is important because many of the lessons of unpredictability apply more broadly beyond self-driving cars to other self-directed, computer controlled moving systems, such as robots.

196 *See, e.g.*, Tom Vanderbilt, *Let the Robot Drive: The Autonomous Car of the Future is Here*, WIRED (Jan. 20, 2012, 3:24 PM), http://www.wired.com/2012/01/ff_autonomouscars (referring to autonomous vehicles as robotic vehicles).

197 The definition of a robot is contested, but there are some common themes. *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CAL. L. REV. 513, 531–32 (2015) (describing the difficulty in defining robotics but providing some basic definitions); Williams, *supra* note 178, at 45 ("Robots are computer controlled cyberphysical systems that perceive their environment using sensors and undertake physical action using actuators to effect change.").

Already researchers in university laboratories and workers in many factories work in close proximity with research or industrial robots.[198] Many expect robots to move out of these specialist settings in upcoming decades, with workers and consumers operating near large autonomous or semi-autonomous robots that assist in work.[199] The important point is that similar issues of predictability about robot behavior and physical harm will become important when people and robots increasingly share the same physical spaces. For instance, if a worker is operating near a large autonomous robot with a moveable arm, that worker would like to have confidence that the robot has detected his presence, and that the robot is not going to suddenly extend its arm in a way that could injure the worker.

Similar to the context of autonomous vehicles, it is important to consider ways to make the internal states of the robot more transparent (e.g., what nearby people has the robot detected?) and future actions more predictable (e.g., is the robot planning to extend its arm?), in order to reduce the risk of accidental physical harm. Thus, many of the lessons and implications discussed in this Article in the context of autonomous vehicles apply equally well to the context of other autonomous moving machines that may become more common, such as robots or unmanned aerial vehicles (i.e., drones).

C. *Law Influencing Vehicle Predictability*

Assuming that autonomous vehicles can be better designed to make their actions more predictable to ordinary people, a related point is, what should the government or legal system do, if anything, to encourage or mandate such changes? It is possible that issues of predictability raised might work themselves out over time, absent any government action, as ordinary users simply become accustomed to the behavior of self-driving cars and such behavior becomes more predictable. But this Article raises these issues prominently, because such a scenario of increased predictability is not necessarily inevitable on its own based upon current trajectories. Currently, different manufacturers are creating driverless vehicles that react somewhat differently from one another in different situations and which have varying capabilities from one another. This variation in behavior and

198 James R. Hagerty, *Meet the New Generation of Robots for Manufacturing*, WALL ST. J. (June 2, 2015, 11:08 PM), http://www.wsj.com/articles/meet-the-new-generation-of-robots-for-manufacturing-1433300884.

199 James E. Young, *How to Manage Robots and People Working Together*, WALL ST. J. (June 2, 2015, 11:10 PM), http://www.wsj.com/articles/how-to-manage-robots-and-people-working-together-1433301051 (describing people and robotics working cooperatively).

capability alone will be, on its own, enough to potentially confuse laypersons attempting to predict the behavior of any one self-driving vehicle that they encounter. This Part will discuss various roles the government might take in fostering more predictability, including standardizing aspects of autonomous behavior and helping to coordinate private sector problem-setting.

1. Government in a Coordinating Role

Perhaps the most promising role that the government might play in encouraging design changes in autonomous vehicles to make them more predictable is through coordination of private sector and public sector efforts.[200] In this role, the government would highlight certain problems that need to be addressed—such as vehicular communication of intentions (i.e., how to communicate to those in nearby proximity to the vehicle where the vehicle intends to move next)—without specifying any particular solutions or performance goals. The benefit of such a coordination role is that the government could focus attention on problems of greatest importance, without venturing beyond areas of institutional competence.[201] One concern is that the development of self-driving vehicles is such a new and constantly evolving technological area, that government agencies may not have the technological expertise to prescribe efforts beyond high-level coordination and focus.

A good example of such government coordination in the context of self-driving cars comes from the 2013 National Highway Traffic Safety Administration (NHTSA) "Preliminary Statement of Policy Concerning Automated Vehicles."[202] In this policy paper, the NHTSA took a leadership role in standardizing self-driving vehicle concepts and terminology.[203] Importantly, in this document, the NHTSA also highlighted certain policy and safety issues that must be addressed by industry in the self-driving car area, including suggestions for training and testing.[204] Although they have not done so yet, the NHTSA could similarly address the issues of movement predictability and external communication highlighted in this Article. Such an official federal government policy focus on problematic issues would go a long way to focusing industry attention on these problems.

---

200 *See* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., FEDERAL AUTOMATED VEHICLES POLICY 22 (2016) (released around the same time as the publication of this Article, and which discusses human-computer interface standards in self-driving vehicles).

201 *See* NHTSA, *supra* note 48, at 6 (highlighting the need for better human factors engineering in vehicle to vehicle communication).

202 *Id.*

203 *Id.*

204 *Id.*

2. Standardizing Self-Driving Car Behavior

Increased standardization of autonomous behavior could help make self-driving cars more predictable. Currently, self-driving vehicles are being developed by multiple different companies and research organizations.[205] While these companies share some broad similarities in their approaches to autonomous driving, at the level of vehicle engineering each has tended to produce vehicles that are somewhat different from one another in both design and capabilities. Vehicles from different organizations tend to have a distinct mix of software, sensors, and mapping reflecting that organization's proprietary research. The result is that vehicles produced by different organizations are likely to react somewhat differently from one another in particular contexts while driving on the road. For instance, a self-driving vehicle developed by Google may approach a crosswalk one particular way given its distinct combination of sensors and software and particular design philosophy, whereas, a vehicle developed by Mercedes may react differently reflecting that organization's unique engineering approach. The result of such variation may be increased unpredictability from the point of view of pedestrians and other lay persons who interact with multiple brands of self-driving vehicles. They may be faced with a range of potentially different autonomous behavior depending upon the source of the vehicle that they happen to encounter.

One approach to increase predictability may be to standardize certain self-driving vehicle behaviors. Such standardization could occur across brands—to ensure that vehicles produced by different manufacturers operate fairly similarly at a broad level. However, such standardization should also occur at the level of common driving contexts that are likely to occur. For instance, a common scenario that has been discussed involves autonomous vehicles approaching crosswalks with pedestrians present. Some sort of standard signaling protocol—for instance—flashing headlights to a pedestrian to indicate that she has been detected or something similar—could be developed for this scenario. If a standard set of best-practice protocols could be developed, these could be implemented across manufacturers. Such standardization of common behaviors would likely improve movement predictability of self-driving cars from the vantage-point of pedestrians, drivers, and other lay persons. Over time, predictability of movement

---

205 More than ten major companies are developing self-driving car technology, along with multiple universities. *See, e.g.*, Danielle Muoio, *10 Companies Making a Bold Bet That They'll Have Self-Driving Cars on the Road by 2020*, TECH INSIDER (Oct. 8, 2015, 11:47 AM), http://www.techinsider.io/google-apple-tesla-race-to-develop-self-driving-cars-by-2020-2015-10.

would increase as pedestrians would become accustomed to a standard set of self-driving car behaviors as they encounter them in the world.

Such standardization often requires a coordinating mechanism, and in some cases, an enforcement mechanism, to actually occur. The government could play a role both in developing best practices for standardization and enforcing those standards. The government is certainly not the only coordinating mechanism that could be used to achieve such standardization—manufacturer or industry groups could work outside of the government or in conjunction with the government—to develop predictable behavioral standards. However, it does seem likely that the government could play a useful coordinating or regulatory role in standardizing autonomous behaviors to make autonomous vehicles more predictable.

3. Direct Legal Influence

Another approach to encouraging design changes in autonomous vehicles to make them more predictable is through direct regulation by an administrative agency. Today, the NHTSA promulgates detailed administrative rules about how to design ordinary (non-autonomous) vehicles to increase safety.[206] These design requirements can be quite detailed and specific. For instance, Federal Motor Vehicle Safety Standard (FMVSS) 101 provides more than five pages of comprehensive rules about the location, visibility, and understandability of vehicle dashboard information.[207] Similar highly detailed government rules exist for nearly every aspect of vehicular design.

One could imagine a series of analogous federal standards designed specifically for the set of unique issues posed by autonomous vehicles. Among these could be design issues related to communicating the intentions of autonomous vehicles in ways that parallel the way human drivers communicate their intentions to others.[208] The code could include a series of design principles specifically mandating that the vehicles meet communication and predictability standards for outside lay audiences. It seems reasonably likely that that the NHTSA or some other federal agency will be involved in regulating issues that are specific to autonomous vehicles that are distinct to those currently faced by non-autonomous vehicles. Although such rules are currently not in place, it seems reasonably likely such federal autonomous vehicle specific regulations will ultimately begin to emerge.

---

206 49 U.S.C. §§ 571, 30111(a) (2012).

207 *See* 49 C.F.R. § 571.101 (2016).

208 For preliminary policy thoughts in this area from the National Highway Traffic Safety Administration see NHTSA, *supra* note 48.

While federal regulation represents one possible mode of involvement, there are reasons to be cautious about such a top down approach. One reason for hesitation is that it is very difficult to understand, at this early stage, how autonomous vehicles will actually unfold once they are on the road in substantial numbers. Currently, fully autonomous vehicles are comparatively rare on the road and largely exist as experimental prototypes. It is very difficult to project forward into a world in which fully autonomous vehicles are more common, representing ten percent or more of the vehicle population. The extreme complexity of the technology, combined with the complexity of multiple autonomous vehicles interacting with one another and human drivers, pedestrians and cyclists, will create scenarios that are today hard to anticipate from a regulatory perspective.

To the extent that such design regulations are promulgated, they should be stated broadly at a high level of abstraction. For one, there is the institutional competence issue. In the case of technological issues, one dominant question is who is in the best position to articulate relevant standards, a government administrative agency or the researchers involved in developing the technology? We are concerned that government agencies may lack the expertise and reaction-speed to produce useful and relevant standards in the face of an emerging, evolving, and iterating technology such as autonomous driving.

Similarly, regulatory rules requiring increased communication for predictability should be promulgated in a functional manner as performance standards. It is not uncommon today among auto design regulations to propose similarly broad rules, where some mandates are stated at a general functionality level in terms of broad standards of performance.[209] With such performance standards, auto-makers are permitted to design their vehicles according to their own specifications and then certify that their vehicle design meets performance standards.[210] Similarly, any regulations related to making autonomous vehicles more expressive for predictability purposes should be specified in terms of performance standards.

4. Indirect Legal Influence

Another way in which the legal system might impact the design of autonomous vehicles to make them more predictable, is not directly through explicit regulation, but indirectly through the tort system. For

---

209 *See, e.g.*, 38 C.F.R. § 17.155 (2016); 40 C.F.R. § 51.351(i) (2016).

210 *See* Stephen P. Wood et al., *The Potential Regulatory Challenges of Increasingly Autonomous Motor Vehicles*, 52 SANTA CLARA L. REV. 1423 (2012).

one, fear of tort liability in accident scenarios resulting from unpredictability and lack of communication—such as those described—might be incentive enough to induce those firms that are developing autonomous vehicles to focus on the issue of unpredictability with more intent.

At the moment, it appears much of the existing self-driving car research effort is focused on solving the fundamental technological challenges in getting autonomous driving technology to work in all weather conditions and physical spaces.[211] For instance, as described, some autonomous vehicle approaches still have difficulty navigating in snow. Thus, there still are central technological issues that remain to be solved through additional research effort.[212] In some ways, the issue of communication and design is likely to be a secondary task once the central challenges of autonomous driving have been mostly solved. Thus, once the technology is fully operational in all conditions, firms may begin focusing more intensely on other issues such as technological design improvements for predictability out of a concern for liability and also a desire to tweak the safety of autonomous vehicles even further.

The tort system might also indirectly affect the technological design through the evolution of case law. Actual accidents and lawsuits might involve scenarios in which the behavior of autonomous vehicles were not predictable to ordinary people. This Article does not aim to do a full analysis of the issues concerning tort liability involved in autonomous vehicle accidents. That analysis has been done ably elsewhere and is beyond the scope of this Article.[213]

However, one final point worth mentioning in tort law: there is likely to be much more data about what happened in an accident with a self-driving car, as compared to a typical car accident today. Such data could serve as detailed evidence as to what happened, and perhaps who was at fault in a tort lawsuit. As described, autonomous vehicles use detailed sensors to collect vast amounts of data about their surroundings. One interesting aspect of this is that when there is an accident, there is a very detailed "black box" record of exactly what happened in both video and data, about the surrounding vehicles, where they were, and what they were doing before and after the accident.

A good example of this comes from another accident involving an autonomous vehicle from Google.[214] That vehicle was hit from behind

---

211 *See, e.g.*, NHTSA, *supra* note 48, at 5–8 (focused primarily on driver safety issues).

212 *Driverless Ford Tackles Snow Problem*, BBC NEWS (Jan. 11, 2016), http://www.bbc.com/news/technology-35280632; Naughton, *supra* note 74.

213 *See*, *e.g.*, Jeffrey K. Gurney, *Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles*, 2013 U. ILL. J.L. TECH. & POL'Y 247 (2013).

214 Chris Isidore, *Injuries in Google Self-Driving Car Accident*, CNN: MONEY (July 17, 2015, 12:04 PM), http://money.cnn.com/2015/07/17/autos/google-self-driving-car-injury-accident.

by another vehicle driven by a person. The autonomous vehicle was not at fault in the accident, but the human driver behind was. Google was able to release a detailed replay of the accident, showing an animation of the accident before, during, and after. Since the lidar and sensors had captured all of the objects around the car for multiple meters and their movements, it became quite clear from the reconstruction of the data, that the autonomous vehicle did not cause the accident.[215] This type of detailed, animated data concerning the immediate movements before, during, and after accidents, which is already naturally captured by the autonomous vehicle's data, will provide a new source of evidence in tort law. The ability to replay and review car self-driving accidents in detail after they have occurred may influence the re-design of self-driving cars for improved predictability.

CONCLUSION

Autonomous vehicles are likely to bring safety benefits. However, one area of concern has largely been overlooked: the ability of lay persons, such as pedestrians or other drivers, to predict the movements of computer-controlled vehicles. This is a problem because today, people rely upon cognitive intuitions about human behavior to avoid accidents with automobiles driven by other people. These same cognitive intuitions may not reliably apply when movement decisions are made, not by people, but by computer systems employing algorithms and sensor data.

This Article had several aims. First, it intended to identify the problem: that the actions of moving autonomous systems may be difficult for lay persons to predict. Autonomous vehicles represent the first example in which lay populations will be operating in close proximity to large moving machines, whose activities are computer-directed (rather than under human control), and that have free range of movement. In coming years, there are likely to be other, similar examples of autonomous movement in non-specialist spaces beyond the autonomous vehicle context, including robots and drones. With such autonomous movement, there is the potential for physical injury. The issues raised in this Article thus apply in these other autonomous contexts (i.e., robots and drones) as well.

This Article also explained the technology underlying autonomous driving with the goal of illustrating why their movements are likely to be

---

215 Hannah Parry, *All Google's Self-Driving Car Crashes Were Caused by Humans, Testers Claim*, DAILY MAIL (Oct. 11, 2015, 11:29 PM), http://www.dailymail.co.uk/news/article-3268421/All-Google-s-self-driving-car-crashes-caused-humans-testers-claim.html#v-4358973660001.

unpredictable in certain high-conflict contexts that are today unproblematic with human driven automobiles. It developed the concepts of "technological opacity" and "theory of mind" to explore why existing cognitive intuitions may not guide lay persons in sharing the same physical space with autonomous vehicles. This Article also explored the way in which improved technological design of autonomous vehicles might make them more communicative in ways that could reduce the risk of accident. Finally, it explored various ways in which the government might play a role in improving predictability, such as in standardizing autonomous vehicle behavior, or fostering technological designs to make the vehicles more communicative about their intentions to lay persons that share their physical space.