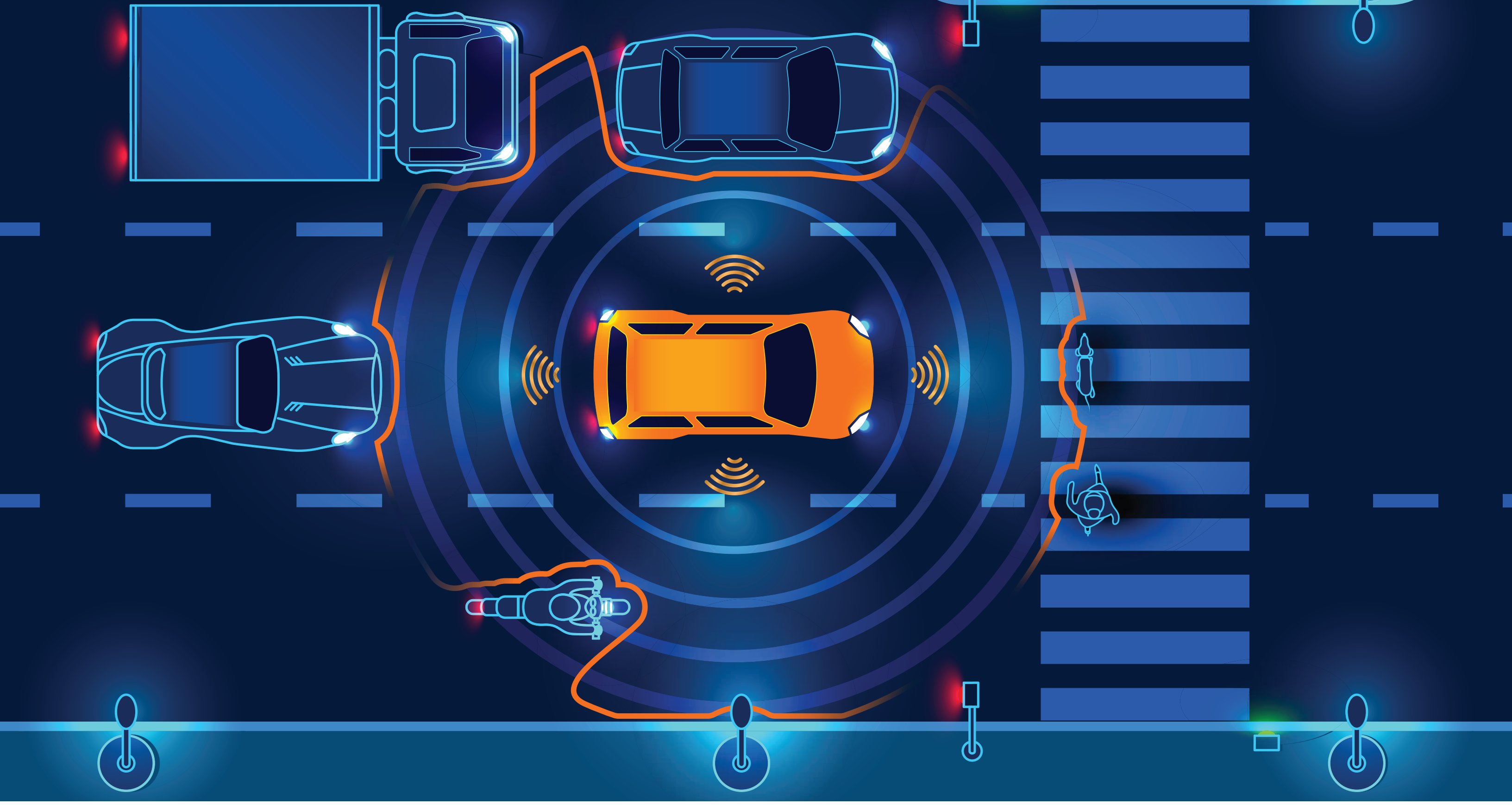


Measuring Automated Vehicle Safety

Forging a Framework

Laura Fraade-Blanar, Marjory S. Blumenthal, James M. Anderson, Nidhi Kalra

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/RR2662

Library of Congress Cataloging-in-Publication Data is available for this publication.
ISBN: 978-1-9774-0164-9

Published by the RAND Corporation, Santa Monica, Calif.

Cover: Courtesy of Andrey Suslov/iStock (Getty Images).

Preface

The safety of automated vehicles (AVs) is intrinsic to their success both in the marketplace and as the kind of transformative innovation that their proponents anticipate. In the summer of 2017, the Uber Advanced Technologies Group approached the RAND Corporation to request help in crafting a framework for measuring AV safety that could aid in public discussion of the issues. This project builds on previous RAND research into AV trends and related public policy. Whereas prior work has addressed issues broadly and analytically, this project has looked more closely at what companies that are key to the evolution of AVs have been doing to foster and evaluate the safety of those vehicles. The report is intended for a broad public audience.

In this report, we develop a framework for measuring safety in AVs that could be used broadly by companies, policymakers, and the public. We considered how to define safety for AVs, how to measure safety for AVs, and how to communicate what is learned or understood about AVs. Given AVs' limited total on-road mileage compared with conventional vehicles, we consider options for proxy measurements—i.e., factors that might be correlated with safety. We also explore how safety measurements could be made in simulation and on closed courses. The closely held nature of AV data limits the details of what is made public or shared between companies and with the government. The report focuses on identifying key concepts and illuminating the kinds of measurements that might be made and communicated.

The research reported here was conducted in two programs. The RAND Science, Technology, and Policy program focuses primarily on the role of scientific development and technological innovation in human behavior, global and regional decisionmaking as it relates to science and technology, and the concurrent effects that science and technology have on policy analysis and policy choices. The program covers such topics as space exploration, information and telecommunication technologies, and nano- and biotechnologies. The RAND Justice Policy Program spans both criminal and civil justice system issues with such topics as public safety, effective policing, police–community relations, drug policy and enforcement, corrections policy, use of technology in law enforcement, tort reform, catastrophe and mass-injury compensation, court resourcing, and insurance regulation. Research in both programs is supported by government agencies, foundations, and the private sector.

RAND Justice, Infrastructure, and Environment (JIE) conducts research and analysis in civil and criminal justice, infrastructure development and financing, environmental policy, transportation planning and technology, immigration and border protection, public and occupational safety, energy policy, science and innovation policy, space, telecommunications, and trends and implications of artificial intelligence and other computational technologies.

Questions or comments about this report should be sent to the project leader, Marjory S. Blumenthal (Marjory_Blumenthal@rand.org). For more information about RAND Science, Technology, and Policy, see www.rand.org/jie/stp or contact the director at stp@rand.org. For more information about RAND Justice Policy, see www.rand.org/jie/justice-policy or contact the director at justice@rand.org.

Contents

Boxes, Figures, and Tables

Boxes

Figures

Tables

Summary

The race to introduce automated vehicles (AVs) to consumers and communities is premised in part on the promise that they will be safer than conventional vehicles (i.e., driven by humans). That race is being undertaken by a diverse, global collection of companies bridging the traditional automotive and technology sectors, including new and established players. Companies have focused on their own innovations in the quest to be the most successful. But the industry, policymaking community, and public can benefit from better ways to understand and discuss the safety implications of AV technology. This report presents a framework to discuss how safety can be measured in a technology- and company-neutral way. Our work focuses on highly automated, complete vehicles operating in the roadway ecosystem. These vehicles are composed of features consistent with Level 4 in the Society for Automotive Engineers taxonomy system accepted by industry and government. The vehicle can perform all driving functions within certain, prespecified conditions.

The meaning of *safety* in regard to AVs is surprisingly unclear—no standard definition exists. In this report, we define *safety* as the eliminating, minimizing, or managing of harm to the public (which can include people, animals, and property, but we focus on harm to people). Following precepts from public health and engineering, we recommend methods to understand progress toward safety.

The public and the policymaking community have an important interest in comparing AV safety with the safety of conventional vehicles. Unfortunately, that comparison is complex because there are different limitations on the breadth and depth of data collected for AVs and for conventional vehicles. Additionally, AVs might have limitations on where and when they can operate. Level 4 AVs operate with full automation within developer-set parameters, which compose the operational design domain (ODD). An ODD can be defined in terms of geography, weather, lighting, roadway markings, previous experience on that specific roadway, and other factors. Comparisons among AVs and between AVs and conventional vehicles must draw their data from the same ODD, otherwise results are likely biased.

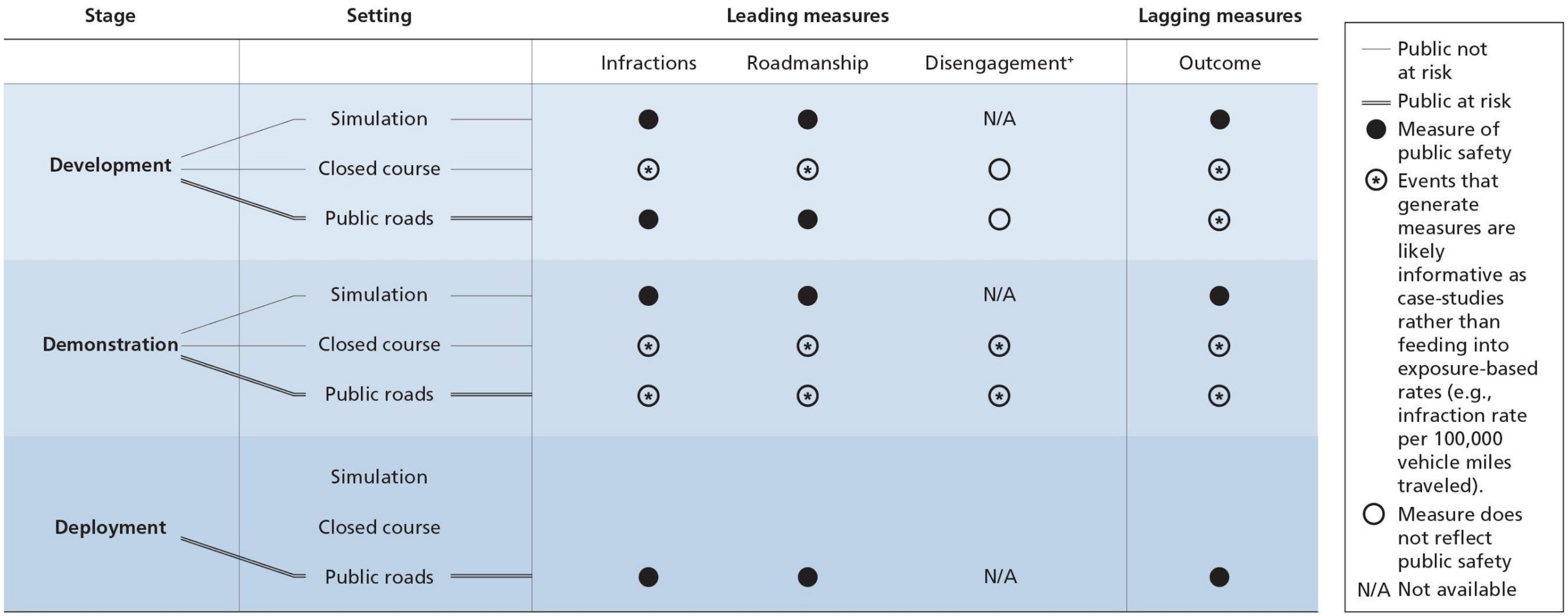
To date, AVs remain in development and are operating only in small numbers in limited situations. That fact underscores the importance of understanding safety through the vehicle model's life stages of development, demonstration, and commercial deployment (Figure S.1). Drawing from conversations with companies, researchers, safety advocates, and policymakers, a research team from the RAND Corporation developed a framework that looks across the settings associated with AV development and demonstration (simulation, closed courses, and public roads) and deployment (on public roads).

The practice that we are calling *demonstration* (while acknowledging that there are other uses of that term) is common in product development. For AVs, it involves ensuring via targeted

testing that what has been developed meets the ODD-specific environment's requirements and user needs, which include being safe. Because AV developers can be expected to do demonstration testing at intervals that mark milestones (other kinds of testing will take place at least regularly but perhaps continuously), those moments provide opportunities for communicating to the public about safety. Testing associated with development prior to demonstration can involve challenging the system as a way to understand its remaining weaknesses. Because finding problems is not intrinsically bad—it can be central to guiding further improvement—most results can reasonably be seen as internally useful to the developer. Demonstration is differentiated here because (1) it provides an opportunity for demonstrating safety at or above a given level (allowing for some level of uncertainty and recognizing that there are limits to what can be demonstrated absent the accumulation of hundreds of millions or more miles driven); (2) its results could be made visible to external parties (something that can differentiate it from routine validation); and (3) it could be undertaken using a protocol that is (semi- or completely) uniform across AVs (building on common attention to industry standards). Such a protocol could involve sets of AV activities in a prespecified sequence and environment that developers call *scenarios*.

The framework (Figure S.1) shows measurement possible in each setting (simulation, closed courses, and public roads with and without a safety driver) at each stage (development, demonstration, and deployment). A black circle or circle around an asterisk indicates measures of public safety. An open circle indicates measures that do not reflect public safety but could be used internally by a company to measure vehicle function and progression. An N/A (not applicable) indicates that the measures cannot be calculated in this setting at this stage. Measures must be valid, feasible, reliable, and non-manipulatable. Measurements can be *leading* (i.e., proxy measures of driving behaviors correlated to safety outcomes) or *lagging* (i.e., actual safety outcomes involving harm). Leading measures could include infractions (failures to follow traffic rules), an integrated measure of driving abilities that we call *roadmanship*, and disengagements (occasions when a person has to take over the control of the vehicle from the automated system). Lagging measures are outcome measures, including crashes, injuries, deaths, and cost of property or vehicle repairs or of medical expenses (for both short and long terms). A third type of measurement, around standards, processes, procedures, and design, is not included because it can be applied at lower levels of vehicle composition (i.e., not for complete vehicles) that evolve constantly, and because these measures relate indirectly to safety through function. With exceptions, each measure can be obtained in each setting at each stage, but the interpretation and utility differs. For example, a measure of infractions on public roads during development might indicate the vehicle's ability to act without legal violations within an environment where the developer has little control beyond ODD specifications. A measure of infractions in closed course testing during demonstration might indicate the vehicle's ability to operate legally within predesignated maneuvers or scenarios, gauging readiness for public deployment.

Figure S.1. Integrated Safety Framework



Stage	Setting	Leading measures			Lagging measures
		Infractions	Roadmanship	Disengagement+	Outcome
Development	Simulation	●	●	N/A	●
	Closed course	⊛	⊛	○	⊛
	Public roads	●	●	○	⊛
Demonstration	Simulation	●	●	N/A	●
	Closed course	⊛	⊛	⊛	⊛
	Public roads	⊛	⊛	⊛	⊛
Deployment	Simulation				
	Closed course				
	Public roads	●	●	N/A	●

─	Public not at risk
═	Public at risk
●	Measure of public safety
⊛	Events that generate measures are likely informative as case-studies rather than feeding into exposure-based rates (e.g., infraction rate per 100,000 vehicle miles traveled).
○	Measure does not reflect public safety
N/A	Not available

+ This column assumed that, in the closed course and public road settings, a safety driver is available (either in the vehicle or remotely). If a safety driver is not present, this entire column would be N/A.

The framework focuses on measures that reflect public safety—i.e., the risks to which the public is exposed. However, measures for which the public is not at risk (the first, second, fourth, and fifth rows in Figure S.1) are obtained in artificial environments where, by definition, the public is not exposed to risk. These rows are included because they represent gateways for development on public roads and deployment, and their measures are determinative of safety, even if they are not themselves situations in which public exposure occurs.

Leading measures are particularly important for AVs because their events happen more frequently than lagging measures. This frequency allows for statistically meaningful, exposure-based comparisons to conventional vehicles to be made at lower levels of exposure (i.e., accumulated mileage). Consequently, an understanding of the AV's safety can be considered earlier in the development and deployment stages. Measures unlikely to accumulate sufficient occurrences in a given stage and setting to permit statistically meaningful measures, and whose events could instead serve as a basis for case studies, are noted in Figure S.1 by a circle around an asterisk (as opposed to a black circle).

Clearer communication about safety between the industry and the public will be critical for public acceptance of AVs. The more consistent the communication around AV safety from industry, the more cohesive and comprehensible the message will be. In addition to embracing shared ways to describe and discuss AV development and safety, clarity is needed on how to think about AV safety (both in absolute terms and in terms that are relative to conventional vehicles), how to measure safety when the AV system continually updates, and how to maximize safety in the face of unknowns.

In addition to the framework, the following recommendations are offered:

- During AV development, regulators and the public should focus their concerns on the safety of the public, not on how development is progressing per se (which is the developer's concern).
- The opportunity to leverage a demonstration stage as a time for communication outside a company about safety (e.g., to policymakers or the public) should be pursued, recognizing the limits to what can be shown absent hundreds of millions or more miles driven and that there is currently no accepted, industrywide approach to demonstration because of variation among companies and the technologies they use. Notwithstanding aversion to sharing information, competitors could report on progress at key demonstration points and adopt common protocols to facilitate fair comparisons.
- Safety events arising before the accumulation of exposure sufficient for statistically meaningful comparisons should be treated as case studies. Information from case studies can contribute to broad learning across the industry and by policymakers and the public. This is happening de facto when an incident is the subject of investigation by government entities; if embraced by industry, it could happen more systematically and more fully.

- Given the potential for broader learning across industry and government, a protocol for information-sharing should be encouraged. It would have to precisely incorporate measures, format, context, frequency, governance, data security, and other factors.
- A taxonomy for common use that facilitates understanding of and communication about operational design domains is needed. A common approach to specifying where, when, and under what circumstances an AV can operate would enable, in particular, inter- and intraorganizational communication and communication with consumers and regulators. It would also facilitate tracking for a given AV of its progress through development and into deployment. Minimal-risk conditions should also be included. Such taxonomies are currently under consideration by industry groups.
- Research is needed around how to measure and communicate AV system safety in an environment wherein the system evolves through frequent updates. AV safety measures must balance reflecting the current system's safety level with recent (and perhaps non-recent) safety records.

Although it can affect AV safety, cybersecurity raises other issues and is not explored in detail. Also not discussed are the special issues associated with automation in commercial vehicles.

Acknowledgments

We are grateful for the many who contributed to this work. Invaluable insight through either individual comments, participation in a workshop, or both was gained through conversations with safety experts in key companies producing automated vehicle technologies, including Baidu, Continental Corporation, Cruise Automation/GM, Intel/Mobileye, NVIDIA, Tesla, Toyota Research Institute, Voyage, and Waymo. We also benefited from discussions with research organizations, including California PATH at the University of California, Berkeley; Exponent; kVA; the Insurance Institute for Highway Safety; and the PEGASUS Project. Conversations with Advocates for Highway and Auto Safety and the National Safety Council contributed additional public safety considerations. Perspective on government thinking was provided by groups of officials from the National Highway Traffic Safety Administration and from the National Transportation Safety Board. Noah Zych and his team at Uber Advanced Technologies Group provided both financial support that made the project possible and insight into their evolving thinking about and expertise in automated vehicle development. Finally, we benefited from the feedback and suggestions provided by colleagues (Liisa Ecola, David Groves, Bradley Martin, Timothy Marler, Lauren Mayer, and RAND summer associate and Massachusetts Institute of Technology graduate research assistant Jeremiah Robertson) during the course of the project. We particularly appreciated the incisive and insightful comments of the reviewers who contributed formally to RAND's quality assurance process (Steven E. Shladover, University of California, Berkeley, and Bonnie L. Triezenberg, RAND Corporation). Responsibility for the content, of course, rests with us and with RAND.

Abbreviations

ADAS	automated driver assistance system
ADS	automated driving system
ADS-DV	ADS-dedicated vehicle
AV	automated vehicle
FARS	Fatality Analysis Reporting System
FMVSS	Federal Motor Vehicle Safety Standards
HAV	highly automated vehicle
ISO	International Organization for Standardization
N/A	not applicable
NASS/CDS	National Automotive Sampling System/Crashworthiness Data System
NHTSA	National Highway Traffic Safety Administration
ODD	operational design domain
OEM	original equipment manufacturer
PEGASUS	Project for the Establishment of Generally Accepted quality criteria, tools and methods as well as Scenarios and Situations for the release of highly-automated driving functions
RSS	Responsibility Sensitive Safety
SAE	Society of Automotive Engineers (now known by acronym alone)
V2X	communication between AVs and infrastructure or other vehicles
VMT	vehicle miles traveled

1. Introduction

The emergence of automated vehicles (AVs), particularly highly automated vehicles (HAVs) that can operate without the expectation that a human would need to intervene to ensure safety, raises practical and policy questions.[1] Safety is of particular concern. It is a prerequisite to trust in AVs and to their widespread deployment. As the U.S. Secretary of Transportation has observed, public and consumer perceptions of safety will drive growth of this technology.[2] Given AVs' differences from conventional cars and the high levels of AV dependence on computer-based technologies, how to think about, measure, and discuss AV safety is not obvious. That is no small irony; proponents of AVs have emphasized their potential to be safer than the conventional vehicles they are expected to displace. In this report, we provide a framework for measuring safety that can be used as a guide for gathering and using data; developing methods; and initiating conversation about HAVs among stakeholders, especially technology developers, policymakers, and the public—consistent with Level 4 in the classification system from the Society for Automotive Engineers.[3]

Context for Contrasting AV and Conventional Automotive Safety

AVs bring at least three challenges to understanding safety. First, they take the use of computer-based technologies—increasingly important to conventional vehicles—to a new level. Computers have always had a black-box, opaque quality to the public, and an AV's dependence on computer-based systems amplifies that opacity. Compounding this is the growing role of machine learning and other aspects of artificial intelligence that provide adaptability to new environments, unpredictability in behavior, and difficulty in tracing decisionmaking.[4] All automobiles are complex systems, built of components and subsystems, but a conventional car's engagement of a human driver makes it more comprehensible to the public than an AV, designed to operate without a human driver's inputs and decisions.[5]

Second, the promise of AVs has brought new players. The latter part of the 20th century saw increasing internationalization of the automotive industry, but the basic structure of original

[1] These vehicles are also referred to as *autonomous*.

[2] Chao, 2018.

[3] The Society for Automotive Engineers is now known as SAE International and referred to as SAE throughout this report. SAE International, undated.

[4] Machine learning systems, for example, evolve as they learn from the data they ingest. This makes predicting the actions of machine learning systems harder to predict than the actions of rule-based systems. See Koopman and Wagner, 2018.

[5] A lack of a sense of control has been associated with the perception of a lack of safety or of an elevated risk. See Möller, Hansson, and Peterson, 2006.

equipment manufacturers [OEMs]) and their suppliers (which can also have their own suppliers) persisted, along with the OEMs themselves (although there was consolidation through mergers and acquisitions). The AV industry is more heterogeneous and dynamic than the conventional automotive industry with which it overlaps. Suppliers of key components or subsystems and developers of complete vehicle systems with roots in the tech sector (semiconductor and computer hardware, software, and services) have entered a space long dominated by companies with manufacturing and mechanical engineering roots. Companies focusing on mobility as a service (e.g., ride-share services) have also entered the fray because of the potential synergy of AVs with new service models. Cross-fertilization of different communities and cultures has emerged, involving both the supply of vehicles for adaptation by new AV-focused entrants and the expansion into AV production by conventional automotive manufactures. Differing company roots, together with employees from a wide range of disciplines, have generated varying—and, at times, diverging—concepts regarding safety.[6]

Third, AVs are one of many areas in which computer-based innovation seems to outpace policy. There are frameworks at the national and state level to foster and establish a lower bound of motor vehicle safety;[7] similar structures are found in other countries, and the international nature of the motor vehicle marketplace guarantees a (mostly) common floor for design, operation, and safety. These range from procedures associated with human drivers (codification of road rules and mastery of those rules and operational skills marked by a driver's license) to compliance with government standards covering vehicular features,[8] many of which anticipate a human driver's involvement.[9] Given compliance with or exemption from applicable standards, AV innovation is supported by the absence of explicit premarket authorization in the United States, although local permission to operate on public roads during development is sometimes required (and has been rescinded following safety events).

Measurement Framework Focuses on System and Ecosystem Levels

Safety can be affected at and by different compositional levels within the AV in the following ways:

- At the most micro end of AV composition (see Figure 1.1), an automated driving system is composed of subcomponents, such as various chips that provide key information-processing functions or camera lenses that contribute to machine vision.

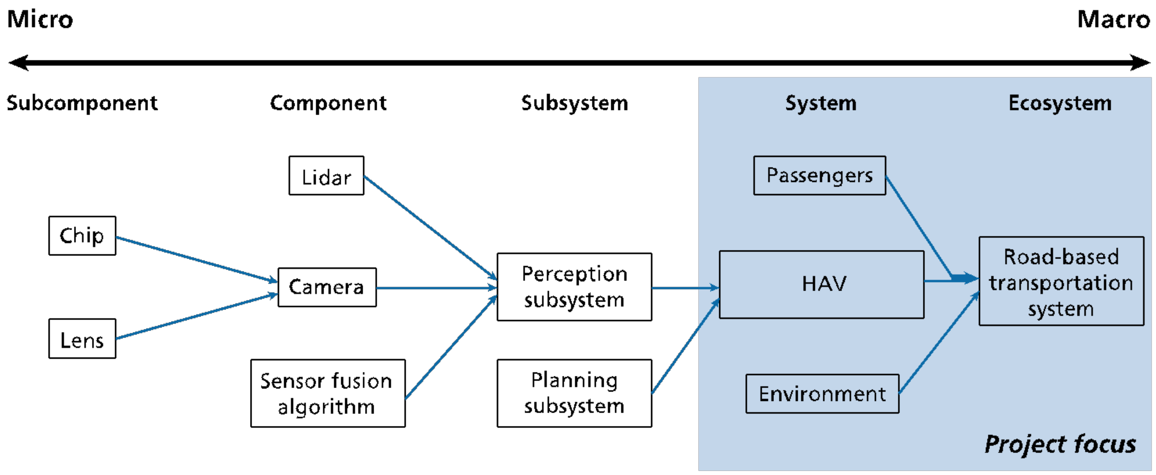
[6] Roose, 2017.

[7] Anderson et al., 2016; Fraade-Blanar and Kalra, 2017.

[8] In the United States, the National Highway Traffic Safety Administration within the Department of Transportation promulgates Federal Motor Vehicle Safety Standards (FMVSS) that implement safety laws and are intended to inhibit and reduce damage from crashes associated with vehicle design, construction, or use. FMVSS are discussed in more detail in Chapter 2.

[9] The expectation of a human driver was the focus of a recent review of FMVSS. See Kim et al., 2016.

Figure 1.1. AV System Spectrum (Notional)



- At the next level are components, such as the cameras themselves, lidar systems that help to measure distance using laser reflection, or algorithms that fuse information from different sensors.
- At the third level are key subsystems, such as those responsible for perception processes and for planning how the AV will move through the environment it perceives.

The macro-level systems of AV composition consist of the vehicle itself, its occupants (themselves representing the complex interaction of cognitive, sensory, and physical functions), and the driving environment. At the highest level is the entire road-based transportation ecosystem, including all vehicles, infrastructure, road users, and attendant legal and medical entities.

The race to perfect AVs involves vendors choosing and developing technologies in specific, idiosyncratic ways—their choices define how they compete. The technologies closer to the micro end of the spectrum are associated with specific vendors and designs and are therefore not agnostic to technology or process. The performance of a lens or camera might indicate how well it sees the world but not the safety-related benefit or loss from its presence or quality of vision. This project addresses safety (of HAVs) at the macro end. Focusing on complete vehicles and their ecosystem makes it possible to discuss concepts independent of specific technology choices.

The potential for AVs to be safer than conventional vehicles has been one of the chief motivators for their development, just as other kinds of robotic systems have been developed to put fewer people at risk of harm in such contexts as manufacturing or mining. As documented in previous research published by the RAND Corporation, safety expectations compare AVs with conventional vehicles driven by people, focusing on vehicle performance or experience rather than components.[10] Comparing these machines at vehicle (or higher) levels fits people's intuition about safety.[11]

Scope, Approach, and Limitations of This Report

This report is intended to foster broader understanding and discussion of AV safety that can aid the public and policymakers in the debate over a new product category expected to reshape economic, social, and community activity.

We developed a framework for AV safety that is neutral in terms of processes, technologies, and vendors and that focuses on SAE Level 4—highly, but not fully, automated vehicles.[12] Level 5, or fully automated, vehicles can perform all driving functions under all conditions that a human could; at Level 4, the driving can be fully automated under specified but not all conditions. We considered different kinds, combinations, and interpretations of evidence for measuring AV safety. Key inputs included review of a variety of literatures (engineering, epidemiological, business, and public policy) and formal exercises and discussions with experts in industry (from the United States and other countries), government, and academia. In those discussions, preliminary ideas were tested; the team's thinking evolved in multiple iterations in response to a steady stream of feedback, new inputs, and discussion.

A principal constraint on the articulation of very specific metrics is the absence of publicly available data.[13] AVs' computer-based technologies collect and analyze vast amounts of data that are proprietary to the vendors and in idiosyncratic formats. Accordingly, the framework that is the heart of this report focuses on underlying measurement concepts and challenges and what could be measured, at least by vendors themselves. Issues relating to data and data-sharing are discussed in Chapter 4 (and in specific contexts in Chapter 3). Although the discussion of safety measurement is relevant to public policy, the project did not examine road safety policy and associated mechanisms in depth.

Three other limitations are (1) the focus on automobiles, recognizing that other kinds of motor vehicles are also becoming automated; (2) the lack of attention to potential key changes in context, such as the shifting percentage of AVs in the fleet and the evolution of communication between vehicles and components of their driving environment; and (3) the acknowledgment but

[10] See Kalra and Paddock, 2016; and Kalra and Groves, 2017.

[11] Smith and Anderson, 2017.

[12] The National Highway Traffic Safety Administration (NHTSA) has embraced the six levels (beginning with no automation) defined by SAE. SAE J3016 designates Level 4 and Level 5 as automated driving systems–dedicated vehicles (ADS-DVs), a "special case" in which "the classification of the ADS and the vehicle are effectively the same." Automated driver assistance systems (ADAS) involve Levels 1–3 and could be designed and combined in different ways in different vehicles. NHTSA, undated-a; SAE Mobilus, 2016; SAE Mobilus, 2018.

[13] As of this writing, the potential to foster and/or pilot data-sharing is under exploration at the Department of Transportation.

not the detailed investigation of cybersecurity as a factor in AV safety. These safety-relevant factors are addressed briefly in Chapter 4.

In this report, we begin in Chapter 2 by considering the challenge of evaluating safety. In Chapter 3, we describe safety metrics as applied to AVs, building to the proposed framework. We conclude in Chapter 4 by putting the framework into a larger context and offering additional recommendations.

2. Safety

For the purposes of this report, we define *safety* as the overall ability of a vehicle to operate without harm to passengers or other road users within the roadway ecosystem. This definition is broadly consistent with other definitions of safety. It focuses on people. Damage to property or infrastructure and injuries to animals are of secondary concern, chiefly of interest because people could have been injured.

Safety exists along a continuum. Safety experts note that both vehicle technology and the larger roadway ecosystem are evolving—what is judged a safe car today might not represent safety tomorrow. As reflected in media coverage and congressional hearings, the average consumer would like to compare the safety of AVs with that of conventional vehicles. Such a comparison might not be straightforward. For example, a conventional vehicle needs a driver, a person whose skill and ability can vary considerably. In the development of a Level 4 AV, there is a person with specialized training in the role of safety driver, effectively a supervisor of the AV, who takes control only in extraordinary circumstances—some of which would challenge human drivers in conventional vehicles.[14] But safety drivers in AVs are not equivalent to human drivers in conventional vehicles.

Safety has been a concern and an objective since the feasibility of AVs was established and commercial development began.[15] In addition to the companies developing AVs, the federal government, governments of states where AV testing takes place, industry consortia and other groups, consultancies focusing on motor vehicle design and development, safety advocates, and academic researchers are among the many actors actively seeking to understand and attend to AV safety. Part of the challenge, as RAND has noted elsewhere,[16] is that a true, comprehensive safety culture remains an aspiration. This chapter addresses some of the conceptual challenges to safety for ground vehicles as a prelude to the more specific discussion of measurement challenges and opportunities in Chapter 3.

Foundational Safety Concepts

There is no consensus definition of *safety* (see Box 2.1)—in general, for transportation, or for AVs. Although the central concept is the lack of harm, variation in how safety is defined reflects

[14] Teoh and Kidd, 2017

[15] The dawn of what has proven to be a race to commercialization is often attributed to the 2004 DARPA Grand Challenge, the goal of which was successful completion of a specified route by an AV. See Defense Advanced Research Projects Agency, 2014.

[16] Ecola et al., 2018.

usage and attitudes concerning adverse outcomes, such as injuries (which could differ in severity) as opposed to death, or injury to an animal or property as opposed to a person.[17]

Box 2.1. Defining Safety—A Sampler

- “The condition of being safe from undergoing or causing hurt, injury, or loss.”[a]
- “Relative freedom from danger, risk, or threat of harm, injury, or loss to personnel and/or property, whether caused deliberately or by accident. See also security.”[b]
- “[T]he problems of transport safety . . . are defined as vulnerability to accidental injury (usually involving at least one vehicle as the instrument causing the injury).”[c]
- “Normally, road safety refers to a range of methods and measures aimed at reducing the risk of accidents. . . . Traffic safety is an indicator of [the] road transport system, as a consequence of the interaction of factors that determine its operation.”[d]
- “In most technical contexts, safety is defined as the antonym of risk. . . . Here, safety is conceived of as a state of low risk: the lower the risk, the higher the safety. This definition is, however, complicated by the fact that ‘risk’ is in itself not a very clear concept.”[e]
- “Safety: absence of unreasonable risk of a mishap resulting in a loss event. Level 4 HAV loss events can include fatalities potentially attributable to HAV design defects or operational faults. For initial HAV deployment, evaluation of what might constitute a ‘reasonable risk’ will be influenced by public policy decisions.”[f]
- The International Organization for Standardization (ISO) defines “safety [as the] absence of unreasonable risk” and “risk [as the] combination of the probability of occurrence of harm and the severity of that harm.”[g]

[a] Merriam-Webster, 2018.
[b] BusinessDictionary, undated.
[c] World Bank, 2002.
[d] Tisca et al., 2016.
[e] Möller, Hansson, and Peterson, 2006.
[f] Koopman and Wagner, 2018.
[g] ISO, 2011.

From an engineering perspective, safety can be considered relative to problems—situations involving vulnerabilities or shortcomings in the system.[18] The risk that a safety problem will occur involves understanding (1) the probability that a situation involving a system vulnerability will be encountered and (2) the severity of the harm incurred. In some cases, steps can be taken to prevent harm; in others, steps can reduce the likelihood of incidents and the severity of harm

[17] The German Federal Ministry of Transport and Digital Infrastructure convened an ethics commission to address issues and attitudes relating to the safety of AVs, and it placed clear priority on protecting people. Federal Ministry of Transport and Digital Infrastructure, 2017.

[18] Hollnagel, Wears, and Braithwaite, 2015. At least in a context such as health care, where humans play an important role, an approach emphasizing what might go wrong can be contrasted with an approach emphasizing what might go right, including how people’s behavior can be helpful.

when they occur. This view of safety can be illustrated with conventional vehicles: Energy-absorbing structures and airbags reduce the severity of harm in the event of a collision; good brakes avoid or reduce the collision energy. Brakes illustrate how automation is affecting conventional vehicles, with automatic emergency braking systems emerging as a faster-acting alternative to human action.[19]

From a public health perspective, safety can be understood as a function of risk—what kinds of harm, with what degrees of severity, occurred relative to the number of people exposed to the potential for harm? For motor vehicles, exposure-based rates of risk have been measured based on population (e.g., per licensed driver or registered vehicle), amount of driving (e.g., vehicle miles traveled [VMT]), and other variables.[20] Even with large amounts of driving data from conventional vehicles, there is still uncertainty about safety.[21]

The Haddon Matrix presents a paradigm used in the field of injury epidemiology to explore such data as injury or crash factors, measures, and interventions (see Table 2.1).[22] The framework organizes factors around the vehicle, the driver, and the environment (including the physical, cultural, and legal environments and the emergency medical system) into pre-crash, crash, or post-crash periods. This is also known as primary prevention, secondary prevention, and tertiary prevention (prevent the crash, minimize injury incidence and severity, and facilitate medical assistance and minimize further injuries, respectively).[23]

As discussed in other RAND work, the novelty of AVs makes the usual approach to measuring safety, which is based on exposure rates, problematic because the accumulation of AV VMT are growing slowly compared with the numbers necessary to draw statistically valid comparisons. Waiting for sufficient exposure before commercial deployment would almost certainly require forgoing the potential benefits of AV use for a very long time.[24]

[19] See NHTSA, undated-b.

[20] The National Safety Council publishes a regular compendium of safety statistics (Injury Facts) that allows comparison of safety across modalities of harm—motor vehicles can be compared with different occupations and situations in homes and communities as sources of harm. National Safety Council, undated-a.

[21] For example, it is known that distracted driving is associated with higher crash risk, but the strength of association between specific types of distractions (e.g., talking on a handheld phone) and crash risk remain unclear. Insurance Institute for Highway Safety, Highway Loss Data Institute, 2018.

[22] *Epidemiology* is the study of the distribution of injury and disease in a population and the reasons for that distribution. Gordis, 2014.

[23] Haddon, 1980; Haddon, 1983; Runyan, 1998.

[24] Kalra and Paddock, 2016; Kalra and Groves, 2017.

Table 2.1. Haddon Matrix

Phase	Human Factors	Vehicle and Equipment Factors	Environmental Factors
Pre-crash	• Information • Attitudes • Impairment • Police enforcement	• Roadworthiness • Lighting • Braking • Handling • Speed management	• Road design and road layout • Speed limits • Pedestrian facilities
Crash	• Use of restraints • Impairment	• Occupant restraints • Other safety devices • Crash-protective design	• Crash-protective roadside objects
Post-crash	• First-aid skills • Access to medics	• Ease of access • Fire risk	• Rescue facilities • Congestion

SOURCE: Peden et al., 2004.

There are a variety of views on how safe AVs should be before commercial deployment. In the United States—and elsewhere, to some degree—the emergence of AVs has been associated at least implicitly with the view that some exposure to risk and uncertainty about this risk must be accepted in the short and medium terms to see the long-term benefit of AVs. Consultations for this project showed how that view, held by many AV developers and by associated industrial and research partners, is not held by the safety advocacy community, which champions clearer communication about risk and more-conservative efforts to at least minimize risk and preferably eliminate it. Given that clear measurement of safety is not likely to be obtained without exposing the public to AV risk on public roads,[25] incremental increases in public AV use are likely to increase our knowledge about safety.[26] The incremental approach can limit risk without stifling the technology. We recognize that the long-term, big-picture view based on confidence in the merits of AVs is shaken every time there is a crash. This report sets out a framework for measuring AV safety across the industry.

Safety Engineering Meets AVs

Safety begins with the design of the vehicle (see Box 2.2), continues with the implementation of that design through the development and production processes, and extends further into the operation of the vehicle.[27] Conventional and automated vehicle production addresses safety explicitly as vehicle design and operation objectives. The development of AVs raises the

[25] Note that testing on public roads is used, in part, to help understand gaps in the design requirements for HAVs—an illustration of the contrast between AVs and conventional vehicles. Koopman and Wagner, 2018.

[26] Fraade-Blanar and Kalra, 2017.

[27] Some AV developers seek to use specialized designs, but because of the huge up-front costs of developing a completely new vehicle platform plus the current regulatory environment, AVs have been produced by adapting and retrofitting conventional vehicles.

question (addressed in Chapter 3) of what should be knowable about safety prior to their commercial deployment.

Ideas about best practices for developing vehicles and related technologies have been codified in voluntary industry-developed standards by SAE and the ISO.[28] These bodies provide guidance for two kinds of safety addressed in both conventional and automated vehicle design and development: functional safety and system safety. In particular, ISO 26262 outlines processes useful for validating the safety of automotive electrical and electronic systems (hardware and software) in the face of a malfunction (a kind of functional safety).[29] It also provides a framework for evaluating associated tools, such as simulators. ISO 26262 starts at the vehicle level, and addresses testing of separate components and their integration. Currently, an AV system that perfectly fulfills its design specifications still could cause or be involved in harm. A pending standard, ISO 21448,[30] often referred to by its name, Safety of the Intended Functionality, aims to address this concern by outlining processes intended to avoid safety problems resulting from unintended behavior (but not malfunction) of sensors, algorithms, and other components or subsystems (a kind of system safety). Achieving functional safety does not imply achievement of system safety, and vice versa. To many experts, compliance with these standards is necessary, but it is not sufficient when it comes to showing evidence of the kind of safety that will be meaningful to the public.[31]

[28] ISO, undated-a.

[29] ISO, 2011, hereafter referred to as ISO 26262.

[30] ISO, undated-b, hereafter referred to as ISO 21448.

[31] The necessary-but-not-sufficient point was emphasized during discussions at an industry workshop convened for this project. Compliance with this kind of standard is also seen as a signal of following best practice, which can be useful to businesses seeking to limit their legal liability.

Box 2.2. Designing Safety into Systems

Designing safety into systems is an art form. Ultimately, the objective is to encourage designers to incorporate safety early in the design process for multiple reasons. First, focusing on safety late in design is expensive and results in delayed programs.[a] Second, safety is different from reliability. Most testing efforts focus on reliability—whether the vehicle performs its intended function most of the time—as opposed to safety—preventing hazards from occurring altogether.[b] Designers focusing on reliability tend to miss safety requirements until late in the development cycle, causing overrun and overbudget programs. One can refine the design and architecture over iterations of the system life-cycle process, improving the overall safety and functional of the vehicle.

Several methods and tools exist that can assist in designing safety into systems. For example, Probability Risk Assessment uses magnitudes and likelihoods of a risk occurring to determine the expected mishaps in any given set of circumstances. Fault Tree Analysis provides a bottom-up approach, starting with the components of a system and gradually identifying how critical certain components are based on their failure chain up to higher levels of the system. System-Theoretic Process Analysis uses a model of the functional control structure of a system to develop safety and security requirements for the system design. Probability Risk Assessments and Fault Tree Analysis focus on hardware-specific failures within a system and can assist subcomponent designers in understanding how their electronics affect the subcomponent. System-Theoretic Process Analysis provides an overall system safety view by analyzing all the components and their interactions within the system, including software, hardware, human, the environment, etc.

[a] Frola and Miller, 1984.
[b] Leveson, ESW, 2008.
NOTE: Box compiled by Jeremiah Robertson.

Federal Motor Vehicle Safety Standards

Accumulated experience with the safety of conventional vehicles is reflected in FMVSS, which specify minimum performance expectations for how vehicles are expected to perform in aspects related to (1) crash avoidance (e.g., rearview mirrors, electronic stability control systems, and tire pressure monitoring systems), (2) crashworthiness (e.g., occupant protection, door locks, and seat belt assemblies), and (3) post-crash survivability (e.g., fuel system integrity, flammability of interior materials).[32] FMVSS do not address AVs (or key automated subsystems, such as for perception or planning) explicitly, and they often assume a human driver.[33] Accordingly, use of conventional vehicle designs ("platforms") can help an AV be compliant with FMVSS,[34] although exemptions can be and have been requested by developers.

[32] Major motor vehicle safety legislation in the 1960s provided for the standards and the establishment of the regulatory entity responsible for their promulgation—NHTSA. NHTSA, undated-e; 49 U.S.C. Chapter 301.

[33] Kim et al., 2016.

[34] Kim et al., 2016.

Implementing FMVSS engages a comprehensive process. Manufacturers certify their compliance with FMVSS, following (often highly detailed) testing procedures established by NHTSA and submitting prescribed documentation.[35] Once the vehicles are deployed (for conventional vehicles, that means available for purchase), NHTSA purchases randomly selected vehicles for testing at facilities it has approved for that purpose.[36] Failing such tests triggers an investigation, which can lead to a recall and/or other actions. Vehicle manufacturers also can (and do) initiate recalls.[37]

Because changes to FMVSS can take years, developers of automated driving systems and the federal government both have sought to reconcile the innovation reflected in vehicle automation with the laws and regulations intended to protect safety. Recent guidance issued by NHTSA calls for voluntary action surrounding automated driving safety, reflecting preferences to avoid both (excessive) regulation and chilling innovation while also protecting safety.[38] That approach sets aside the possibility of the kind of regulation that a heterogeneous industry can agree to as broadly beneficial.[39] Deep consideration of policy mechanisms for motor vehicle safety is beyond the scope of this report.

Special Aspects of Level 4 AVs

As with other levels of automated driving systems, the development and testing of those at Level 4 has involved safety drivers in the vehicle, available remotely, or both. Safety drivers are intended to act as a backup to computer-based systems that are recognized as imperfect, especially during development.[40] Safety drivers can take over when the computer-based system appears unable to deal with circumstances or when the driver believes that the system is not responding adequately to those circumstances. The process of taking over is called *disengagement*. Issues with counting disengagements (something required of those testing AVs on public roads in California) are discussed in Chapter 3.

At the far end of the vehicle automation spectrum, developers associate Level 4 automated driving systems or AVs with an operational design domain (ODD) in which the vehicle can

[35] NHTSA, undated-f; NHTSA, undated-g.

[36] Indicative of the level of activity, monthly compliance report summaries are available online. NHTSA, 2018a. NHTSA Office of Vehicle Safety Compliance Contract Compliance Test Laboratories are listed online. NHTSA, undated-d.

[37] NHTSA, 2017a.

[38] NHTSA, 2017b.

[39] Although uncommon, there are occasions when industry welcomes regulation, as illustrated by the experience with feminine hygiene products when toxic shock syndrome emerged. Miller et al., 2017.

[40] Safety drivers provide what engineers would call a human in the loop; the *loop* refers to the control system of the AV.

safely operate.[41] The ODD could be defined by geography, time of day, weather, or other factors to varying degrees of detail. Individual Level 4 AVs will be designed to provide different kinds of service in different environments, each with its own ODD restrictions. Automated shuttles, for example, which keep to simple and fixed routes and sometimes travel at low speeds, tend to have comparatively simple ODDs. A minimum ODD for a low-speed urban shuttle service will be completely different from a minimum ODD for a long-haul interstate trucking service.[42] The expectation has been that progress in an AV's development enables ODD expansion; showing safe operation in a widening variety of circumstances is key to operating in a correspondingly broader ODD and, at least in principle, eventually without any domain constraints—a Level 5 vehicle should be able to operate anywhere that a conventional vehicle can.[43] Because Level 4 AVs are not expected to operate anywhere and anytime, their ODDs and associated limitations would have to be known to owners and users. (This might be another reason to expect more fleet ownership than consumer ownership).

[41] NHTSA (2017b) outlines the role of ODDs. The industry is guided in more detail by SAE J3016 (SAE Mobilus, 2018).

[42] SAE Mobilus, 2018.

[43] One pair of researchers has observed that developers might prefer Level 4 to Level 5 in order to avoid having to handle all possible scenarios. Koopman and Wagner, 2018.

3. Measuring Automated Vehicle Safety

Measures of conventional vehicle safety have increased in complexity over time with the rise of event data recorders and in-vehicle monitoring. Traditional measures of police-reported crash rates, generally per licensed driver, are now joined by measures of inattention, ADAS activation, etc.[44] But measuring safety in AVs differs from measuring safety in conventional vehicles: The integration of the driver and vehicle invalidates or changes how traditional measures can be used and gives rise to new needs, all complicated by the new array of technology, implementation approaches, data availability, and business models. Additionally, measurement must be appropriate to automation level and ODD.

Our framework to conceptualize, contextualize, and measure AV safety contains the following three frames:

1. **Settings:** contexts that give rise to safety measures
2. **Stages:** the life stages of AV models during which these measures can be generated
3. **Measures:** the meaning of new and traditional measures obtained in each setting as AVs move through each stage.

The following subsections examine each frame and associated measurement issues, building to the overall framework. Because of the ways the issues interact and intersect, sometimes a given issue might be addressed from different perspectives on the way to the culminating framework.

Frame 1. Settings

Settings refer to the situational contexts in which the AV operates. At the vehicle level, there are four settings for measuring safety (Table 3.1): computer-based simulation, closed courses, public roads with a safety driver present or remotely available, and public roads without a safety driver. Each setting varies by the extent to which the public (defined as individuals beyond those employed in producing and testing the vehicle) is exposed to risk and how this risk is controlled. One method to control risk is having what engineers call a human in the loop—a person is involved in influencing safety in a situation. In the case of an on-site or remote safety driver, the person would serve as a fallback during development but not post-deployment.

[44] ADAS covers a wide range of features, from automatic lights and lane-departure warnings to blind-spot monitoring and adaptive cruise control, among others. These depend on inputs from sensors (e.g., cameras).

Table 3.1. Settings Frame

Setting	Safety Considerations	Extent to Which Public Is Exposed	How Risk Is Controlled
Artificial settings: simulation[a]	There is no risk because automated driving is not occurring in a real environment	Not at risk	• No risk exists. The environment is entirely created and controlled by a human, and only artificial humans, animals, and property can be harmed.
Artificial settings: closed course	Driving poses risks to the closed-course property and to the employees of developers or course operators	Not at risk	• Humans might be in the loop as in-vehicle or remote safety drivers • Environment at least semi-controlled
Public roads: safety driver (present or remote)	Driving poses risks to trained drivers, potential passengers, and other road users	At risk	• Human in the loop • Environment is not controlled
Public roads: no safety driver	Driving poses risks to vehicle passengers and other road users. There is no safety operator.	At risk (public includes those inside the vehicle)	• No human in the loop • Environment is not controlled

[a] In this report, simulation refers to computer-based systems that feature no human interaction with the AV hardware or software during the simulation. However, conclusions could be extended to simulator-based driving simulations involving humans where actions such as AV-to-human handoffs are explored. Doing so would relate to the development process rather than anticipated deployment for Level 4 systems, which are intended to obviate such hand-offs.

Artificial Setting

The artificial setting consists of simulation and closed courses. Although these differ widely in their mechanics, both environments are heavily or completely controlled, largely contrived, and pose no risk to the public. In simulations, only simulated humans, animals, and property are involved. There are no safety risks because automated driving is not occurring in real life. On a closed course, automated driving might pose risks to course property, employees of developers, and course operators but not to the general public. Within these settings, developers can safely test dangerous scenarios (e.g., to see whether the automated driving system can discern and respond appropriately to a ball, a small dog, or a toddler crossing the street), using simulated people and animals in simulations and dummies in closed courses. Validity refers to the ability of a test, scenario outcome, or other measure to determine what is safe and what is unsafe.[45] For the simulation setting, validity depends on the correspondence between simulated and actual AV performance. It has two aspects:

[45] Gordis, 2014.

1. **Internal validity:**[46] How good is the simulation, beginning with the quality of the software (e.g., speed, amount of bugs, faithfulness to preset protocol and best practices)?
2. **External validity:**[47] Is simulated AV performance consistent with real-world performance under the same and other parameters (e.g., if the simulation assumes full sun at a 90-degree angle, does the AV system perform similarly in the real world in full sun at a 90-degree angle, and then at an 89-degree angle)? Using simulations requires assumptions. Simulations can be inaccurate or overly simplified versions of the real world. External validity is very similar to the engineering concept of fidelity.[48]

The practical inability to compare internal validity across types of simulations and simulators poses a problem for external validity. As discussed later, competition among AV developers, varying approaches to simulation, and a lack of AV simulation standards constrain comparisons of different AVs.

Simulations also might suffer from *confirmation bias*, the interpretation of measures as confirmation of existing safety levels (see the later section, "Frame 3: Measures"). The bromide that "all simulations are doomed to succeed" captures this bias.[49] Too little rigor, detail, or range in AV simulation, coupled with insufficient analysis of results or their interpretation, can cause a simulation to appear to confirm safety levels while failing to connect simulation to real-world performance. This bias can exist both in simulations of general driving situations (where the outcome is unknown) and in the re-creation of key scenarios identified as challenging by on-road driving (where one is instead trying to understand inputs and explain processing mechanics).[50] Confirmation bias is combated by continual revision of safety measures based on new information and by creative interrogation of results to explore alternative explanations.

Safety flaws can develop at the seams that exist between subsystems, notably as the result of misalignment of each subsystem's purpose and miscommunication of constraints, but not be detected in simulation.[51] Sensor simulation remains imperfect because of the difficulty of simulating the details of the electromagnetic wave propagation that determines the performance of video, lidar, or radar. Hardware and sensors can be added into more-sophisticated simulations, but this artificial setting remains unable to reflect the ability of all the subsystems and their components to work together in harmony in real world environments.[52]

[46] Gordis, 2014.

[47] Gordis, 2014; Hartas, 2015.

[48] Koopman and Wagner, 2018.

[49] Grim et al., 2013; Mynatt, Doherty, and Tweney, 1977.

[50] Grim et al., 2013.

[51] Leveson, 2011.

[52] Koopman and Wagner, 2018.

Another artificial setting, closed-course proving grounds, is closer to the real world than simulations because the vehicle subsystems are fully integrated into an AV system and operated in the physical environment, although environments could be limited and/or highly controlled.[53] Many closed courses exist around the world—some for exclusive use by a given developer or OEM, others available on a membership basis or other paid use. In 2017, the U.S. government "designated 10 proving ground pilot sites to encourage testing and information sharing around automated vehicle technologies," with attention to safety among the criteria for selection.[54] Closed courses allow for quasi–real-world tests of an AV system's design limits and of what occurs when the system exceeds those limits.[55]

Closed courses are subject to the same external validity concerns as simulation. There is a limit to the amount of chaos and unpredictability present in artificial environments.[56] In simulations, the limit reflects computational sophistication and developer creativity; in closed courses, the limit reflects facility capabilities and safety concerns. Intra- and interdeveloper variation in safety driver behavior and variation in closed courses make it difficult to compare results achieved at different sites. Similarly, simulations can vary widely by developer, and comparing results achieved in one simulation to those achieved in another will present difficulties.

Of course, these concerns do not invalidate the immense utility of simulation and closed courses, which are foundational to AV development. Rather, the outlined concerns argue for caution in the interpretation of safety measures from the artificial setting beyond what is supported by the setting.

Public Road Setting With and Without a Safety Driver

Measuring AV safety on public roads involves trade-offs. Because AVs are intended for public road use, safety measures obtained in this setting present fewer concerns around internal or external validity. However, operating on public roads with and without a safety driver involves exposing the public to risk. Also, companies choose when and where the AV begins its travels during development, but, beyond specifying an ODD, they have little control once the vehicle is deployed. Environmental characteristics can change suddenly (e.g., shifts in weather, road construction, unexpected congestion, unexpected behavior of other road users).[57] Additionally, because of the natural chaos of the roadway ecosystem and the variety of factors affecting AV performance, exact repeatability of on-road situations (sometimes modeled in scenarios used for testing) is unlikely. The ability to respond to rare or unusual circumstances,

[53] Maddox, 2018.

[54] U.S. Department of Transportation, 2017.

[55] Maddox, 2018.

[56] Koopman and Wagner, 2018.

[57] There are cases of highly restricted ODDs, such as campus shuttles with very simple routes, but these represent only one specific kind of business model.

discover situations that fall within the ODD but in which the AV cannot function, and evaluate AV system performance after upgrades can be explored on public roads. Driving in closed courses and simulation provide additional settings to explore and circle back to test solutions to identified problems. Some public road settings will be easier than others for AVs to manage because of weather, road conditions, etc.

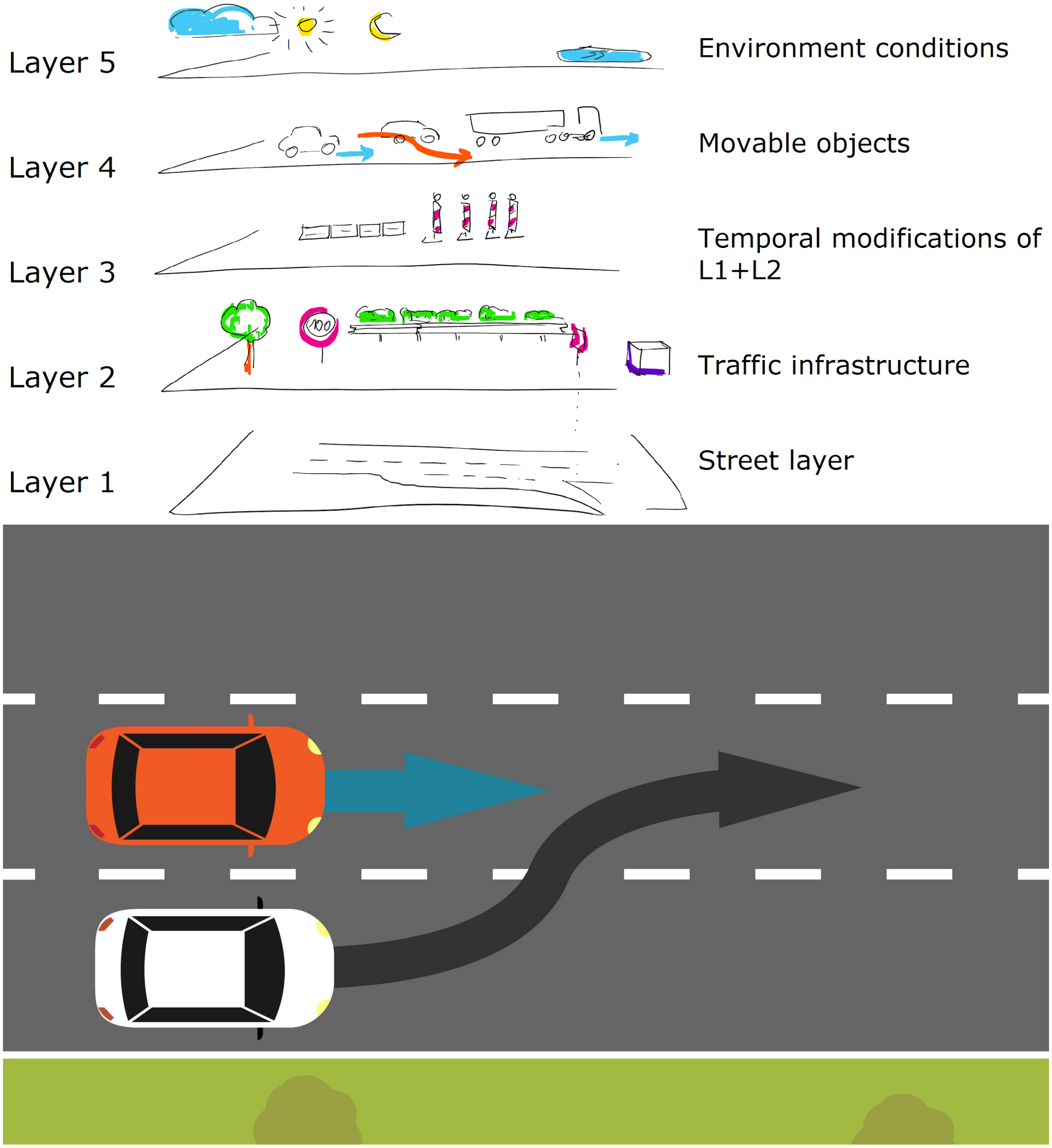
For AVs in the public road setting with a safety driver, a human mediates risk (albeit imperfectly). This mediation is removed when the safety driver is removed, whether in development, demonstration, or commercial deployment. When passengers are present in the vehicle without a safety driver, a remote operator to assist with non–safety-critical issues could be available.

A Discussion of the Setting Frame

During the development and demonstration stages (discussed in the next section, "Frame 2. Stages"), simulations, closed-course driving, and (to an extent) public-road driving of AVs are based on scenarios. *Scenarios* involve a sequence of behavioral competencies in a specified driving environment that exercises and challenges the capabilities of the automated driving system. The level of detail and how precisely it needs to be executed (Figure 3.1) depend on the technical questions it poses.[58] Typical scenario development and implementation is shown in Figure 3.2. Scenarios used in simulation and closed courses are based on conventional vehicle experiences gleaned from sources including naturalistic data (sometimes collected by the research community) and developer-collected, on-road data (loop 1 in the figure). The AV could subsequently enact the scenario on public roads, although the natural unpredictability of public roads means that exact enactment is not guaranteed. Public-road performance and the environments that the AV encounters are fed into simulations and closed-course scenarios, generating an ongoing cycle of gathering and testing scenario-specific safety performance data (loop 2 in the figure). Loop 2 is bidirectional; simulation, closed-course, and on-road experience do not necessarily occur in strict order. Results of driving on closed courses and public roads can be used to improve simulation or generate new simulated scenarios and vice versa. Loop 2 is necessary to check implementation because situations that challenge a human driver might not challenge an AV, and vice versa.

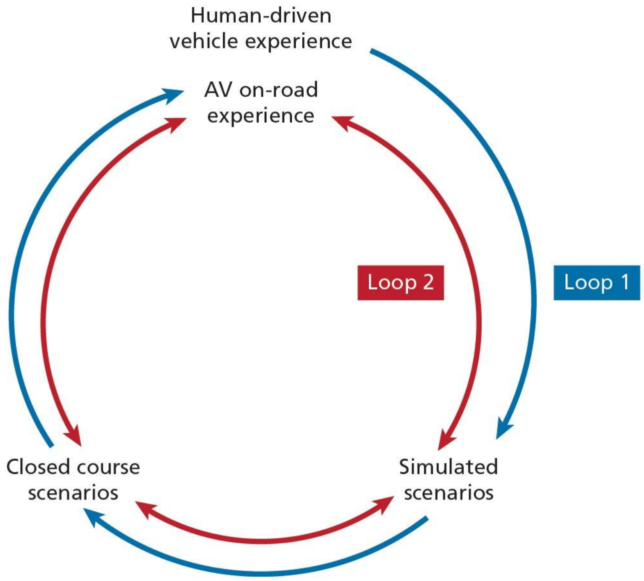
[58] For examples of scenarios, see NHTSA, 2007, and Waymo, 2018.

Figure 3.1. Scenario Layers and Illustrative Cut-In Scenario



SOURCE: Provided to RAND by PEGASUS-Project, Institute for Automotive Engineering of RWTH Aachen University and Institute of Control Engineering of Technische Universität Braunschweig

Figure 3.2. Automated Vehicle Scenario Development and Use



However, these loops risk myopia without the following:

- procedures that introduce faults (a process referred to as fault injection[59]) into the AV system hardware, software subsystems, or communication between subsystems to test robustness
- exploration into rare events not hitherto encountered by the AV[60]
- some other method to challenge AV performance beyond what has been previously demonstrated.

The AV will be prepared only for what it has seen. To be ready for deployment, the AV must be ready for the challenges it will encounter, not just those it has already encountered.

During development and demonstration (discussed in the next section, “Frame 2. Stages”), measurement in simulations, closed-course tracks, and on-road driving with or without a safety driver will always be incomplete. The degree to which one scenario in simulation and closed courses represents all variations in the real world is limited. The range of scenarios explored can never cover every situation. The number of variations in weather, traffic patterns and road users, and so on that are extant on U.S. roadways is finite, but it is sufficiently large to be effectively infinite. The number of miles driven on public roads is limited in this stage. Even the most comprehensively simulated system (including being test-driven in closed courses and on public roads) will still experience surprises because a rare event or surprise might occur less frequently

[59] Jha et al., 2018.

[60] Koopman, 2018; Koopman and Wagner, 2018; Maddox, 2018.

than the amount of exposure experienced during the development and demonstration stages.[61] For example, if a certain type of rare event occurs only once every 10 million miles, encountering such an event even once during development and demonstration is unlikely.[62] Consequently, the ability of the AV to recognize that it is encountering a challenge beyond its experience and capabilities (i.e., outside its intended ODD) and devolve gracefully to a minimal-risk condition must be a key concern.[63] Development includes provision for that process.[64]

Whether the public is exposed to risk and how that risk is controlled varies in different settings—artificial or public roads. Measures of safety gathered in artificial settings allow developers to explore and perfect AV behavior in new scenarios without exposing the public to risk. On public roads, the roadway becomes a living laboratory, and other road users become involved in a study that they did not consent to take part in and cannot opt out of. Because the overall goal is to consider whether the vehicle is safe in terms of the public, how we use measures to find evidence of safety differs between settings.

Frame 2. Stages

Innovation typically contains a development stage, where the product is created and refined, and a deployment stage, where the product is released to the public (Figure 3.3). Sometimes, particularly for software-based technology, refinement and revisions (some characterized as upgrades) continue during deployment. For AVs, safety in development and deployment is evaluated with reference to the public—either occupants of the AV or other road users. Between an AV's development and deployment is a safety demonstration stage (hereafter referred to as simply *demonstration*). Demonstration activities can establish the automated driving system's safety level or provide evidence that it is above an established minimum.[65] The transition to deployment implies achieving a threshold level of safety for the general public, including the consumer. It is anticipated that AV systems will move through these stages in a fashion that is iterative, nonlinear, and bidirectional, as models are deployed commercially while developers work on the next upgrade or system model.

[61] The distribution of the probability of events over time could have what statisticians call a heavy tail. Heavy-tailed distributions have a large probability of getting large values (Wolfram Language and System Documentation Center, undated). For AVs, this means that there will be very rare events within the distribution of the probability of events. An example of a heavy-tailed distribution is annual income in the United States; most adults are within a narrow range but a few individuals are very high earners.

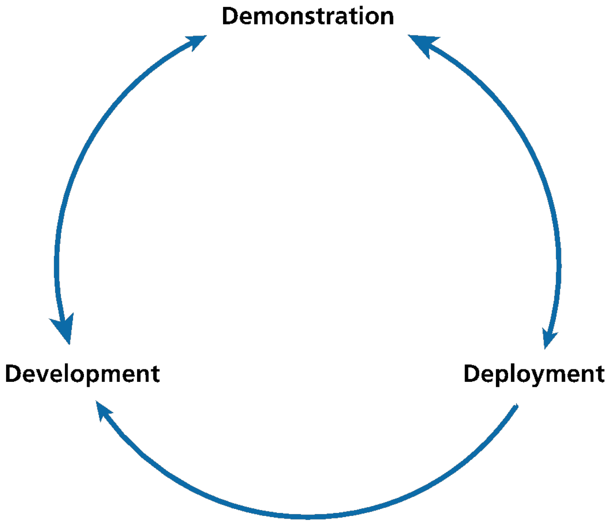
[62] Koopman, 2018.

[63] Koopman and Wagner, 2018.

[64] SAE Mobilus (2018) outlines the need for "fallback performance" and "minimum risk achievement" in the event that there is an "out of operational design domain (out-of-ODD) condition."

[65] Koopman and Wagner, 2018.

Figure 3.3. Stages Frame



Development

Development covers the period from inception until demonstration. The developer's safety goal in this stage is to establish and improve the automated driving system. This includes the following:

- identifying the situational parameters of its intended ODD
- spotting outstanding gaps in system performance
- ensuring that an AV reverts to a minimal-risk condition when the vehicle exceeds its ODD[66]
- avoiding creation of new safety hazards[67]
- systematically reducing the overall uncertainties implicit in each setting with complementary testing in other settings.[68]

System performance at these activities could be assessed during routine verification and validation associated with system engineering. At the development stage, safety refers not to how risk is controlled or whether the AV's development is progressing in a certain path; those reflect the developer's design and process choices. Rather, safety refers to whether the vehicle's behavior creates risk for the public.

During development, safety measures can be generated from scenarios in artificial or public road settings. Many developmental pathways exist, from extensive testing in a private campus

[66] SAE Mobilus (2018) outlines such fallback procedures, along with failure mitigation strategies, as already noted.

[67] NHTSA, 2017b.

[68] Koopman and Wagner, 2018.

system using a rideshare model to running key software in the background (i.e., without affecting the operation) of in-use, conventional vehicles.

Demonstration

The demonstration stage is an exhibition period of sorts—developers demonstrate that the vehicle acts safely (in part, by showing that it does what it is supposed to do and does not do what it is not supposed to do).[69] It can also be a period to obtain data for a Voluntary Safety Self-Assessment (invited by the U.S. Department of Transportation)[70]—data to make the case for the safety of the automated driving system.[71] Demonstration could be conceived as a developer-declared period with some level of oversight or transparency, set at the time of the developer's choosing and consistent with the stated ODD. Although measurements obtained during demonstration can be differentiated from those collected in development and deployment, demonstration testing could be repeated throughout the development-to-deployment life cycle. Having a demonstration period, especially one after which safety measures are made public, allows for more latitude in the development period, during which new challenges and approaches can be broached without fear of undermining the AV's perceived safety level.

Demonstration is distinct from development in the following ways for the following issues, with stage goals being the most important:

- AV system shortfalls
    - In development, identifying AV system shortfalls indicates where the developer can improve.
    - In demonstration, AV system shortfalls could indicate the AV remains unprepared for commercial deployment (or for moving on to some next level of development en route to deployment).
- AV system status
    - In development, the AV system is in flux.
    - In demonstration, the system is relatively stable.
- ODD
    - In development, the system's robustness and reaction to new environments (e.g., ODDs) can be explored and performance tested.
    - In demonstration, all activities occur within stated ODDs. As an exception, the AV could encounter situations outside its ODD in order to assess its ability to detect this and respond safely.

[69] Hollnagel, Wears, and Braithwaite, 2015.

[70] The Voluntary Safety Self-Assessment, created by NHTSA (2017b), provides summary information about safety level to the public and a template for summary information (2017c).

[71] NHTSA, 2017b.

- Stage goals
    - During development, the goal is to become able to show that the AV does not present a danger to the general public—especially people, but also animals and property outside the vehicle.
    - During demonstration, the goal is to show publicly that the AV is safe for commercial deployment or further development by demonstrating safety at or above a given level (allowing for some level of uncertainty and recognizing that there are limits to what can be demonstrated absent the accumulation of hundreds of millions or more miles driven), robustness against surprises, and ability to detect and safely transition in and out of its ODD according to an informal, semiformal, or formal and uniform protocol.

Verification and validation occur routinely during development (Box 3.1) and inform demonstration, which is itself a kind of validation.[72] In verification, the developer examines the AV system to ensure that it performs to specifications or design requirements.[73] In validation, the AV is examined to ensure it performs in ways that directly meet consumer needs safely within its ODD.[74]

Measures obtained during demonstration using simulation could inform conclusions on safety. High-fidelity (excellent external validity) simulations could use a prescribed set of scenarios matched to ODD,[75] such as those under development by the German collaboration, PEGASUS.[76] But demonstrating by simulator presents challenges. As noted previously, simulations can be inaccurate or simplified representations of the world, failing to account for germane vehicle, sensor, or environmental characteristics.[77] Additionally, the lack of current standards around simulator quality makes a uniformly implementable, simulator-based demonstration tricky. Because companies invest considerable resources into their simulation systems, which are specialized to their particular AV system, switching between simulators

[72] ISO 26262 calls for a "V"-model progression through verification to validation. A loose comparison for the contrast between routine validation and demonstration could be made to the formative and summative assessments used in education.

[73] ISO, International Electrotechnical Commission, and Institute of Electrical and Electronics Engineers, 2010. As noted earlier, the use of machine learning in highly automated driving systems can make it hard to understand the full requirements picture.

[74] An evolving counterpart to ISO 26262 is ISO 21448, which is intended to support the goal of designing safety into the vehicle.

[75] Koopman and Wagner, 2018.

[76] PEGASUS stands for "Project for the Establishment of Generally Accepted quality criteria, tools and methods as well as Scenarios and Situations for the release of highly-automated driving functions." The organization has engaged technical personnel from companies and research institutions to develop "a generally accepted and standardized procedure, for the testing and approval of automated driving functions." PEGASUS Research Project, homepage, undated-a; PEGASUS Research Project, undated-b.

[77] Koopman and Wagner, 2018.

could be onerous for developers and could introduce new threats to internal and external validity and reliability. Exploiting the expectation that companies will teach to the test, having common scenarios, simulators, or related standards in the demonstration period could present an opportunity to ensure parity earlier in the development process. Such benefits motivated the PEGASUS Project and other efforts to promote common approaches to testing. Anecdotally, competitive pressures might limit sharing; other aspects of sharing are discussed in Chapter 4.[78]

Measures obtained during demonstration in another artificial setting, closed courses, can assess behavioral competency:[79] the ability of the AV to operate in its ODD.[80] Behavioral competency includes both normal driving conditions and abnormal but foreseeable driving conditions, such as preforming a high-speed merge on the freeway.[81] However, including such challenging conditions is not sufficient to provide assurance of safety.

There is no standard protocol to assess behavioral competency. As an illustrative option drawing from work by the California PATH Program,[82] a manufacturer generates a safety plan including ODDs and behavioral competencies. If acceptable, a third-party tester decides on a set of test cases to be conducted in a closed course. If the performance of the vehicle is sufficient, the tester or state DMV would conduct various maneuvers on public roads.[83] Alternatively, PEGASUS proposed that critical scenarios used in simulation could also be used on closed courses and public roads.[84] Such an approach has the benefit of allowing lessons learned on the closed course to be incorporated into the simulator, gradually increasing the fidelity of the simulations.

The use of closed courses for demonstration presents similar concerns about validity and reliability as found with simulations. Regarding reliability, closed courses offer a range of environments. A tension exists between providing uniform implementation of a demonstration protocol and each closed course developing specialized offerings. Additionally, the range of testable scenarios is limited by the capabilities of the closed courses. There are some essential maneuvers too complex or too dangerous to test (e.g., presence of roadwork with dense traffic or high-speed passing in heavy traffic with a mixed vehicle fleet).

[78] For example, in 2018, the company Vires transferred some of the work it had stewarded on open standards relating to scenarios and simulations to the Association for Standardization of Automation and Measuring Systems to "ensure an independent further development and long-lasting maintenance of the standard." Association for Standardization of Automation and Measuring Systems, undated.

[79] California Code of Regulations, undated; Nowakowski et al., 2015.

[80] NHTSA, 2017b; NHTSA, 2016b.

[81] Nowakowski et al., 2015.

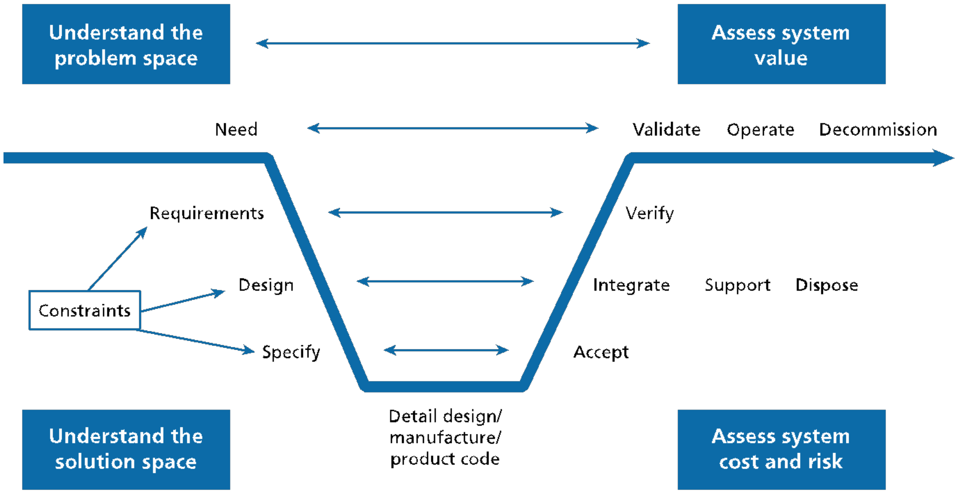
[82] Nowakowski et al., 2016; Nowakowski et al., 2015; University of California PATH Program, 2016.

[83] University of California PATH Program, 2016.

[84] Amersbach and Winner, 2017.

Box 3.1. Functional Safety and Automated Vehicles

Although traditional engineering validation and functional safety focus on different aspects of the system design life cycle, their connection is important. By focusing efforts on designing safety into the system early, validation can focus on system robustness.[a] The traditional approach to systems engineering invokes the V model, so termed because it is typically drawn in the shape of a V. On the left side of the V, the model starts with development of a concept of operations, definition of requirements, development of an architecture, and the building of the system. Heading up the right side of the V, the model ends with integration of the system (if need be), testing, verification and validation of performance, and operation and maintenance of the system until its end of life.[b]



Understand the problem space
Assess system value
Need
Validate
Operate
Decommission
Requirements
Verify
Constraints
Design
Integrate
Support
Dispose
Specify
Accept
Understand the solution space
Detail design/ manufacture/ product code
Assess system cost and risk

SOURCE: International Council On Systems Engineering, *Systems Engineering Handbook,* Version 3.1, August 2007, pp. 3.3–3.8.

By designing safety early on the left side of the V, the right side actions can focus traditional validation efforts on the whole system and in-situ environmental parameters instead of individual components and their inputs and outputs. This methodology will find significant errors during testing that a team would otherwise miss because of the time constraints on testing. However, the traditional V model approach to systems engineering is predicated on the notion that the requirements are known. With automated systems, there are a plethora of unorthodox road settings and environmental hazards that can be examined in the safety demonstration stage to assist in validating the functional safety of the subsystems, completing the right side of the V model.

A few methods help developers handle the complexities of automated systems. One is to limit the ODD in such a way that the safety envelope (a set of limits and conditions under which the system must operate, as discussed in the section of this report titled “Measure Category 2: Leading Measures”) can be expanded and confidence can be built over time. Additionally, developers can use the levers associated with ISO 26262 to assign safety requirements to a high automotive safety integrity level while the functional requirements are allocated to a low automotive safety integrity level. This ensures that all subcomponents are focusing safety requirements around significant mishaps that must be avoided while the lower-level functions can perform as intended beyond the safety constraints.

By appropriately defining requirements and performing specific forms of verification and validation during the design and development process, AV developers can utilize a phased deployment that will build trust with the public while ensuring safety is maintained. While traditional validation through the V model worked well for systems with mostly hardware, modern automation systems are complex and require new forms of validation (the *demonstration stage*) that provide safety assurances beyond individual components. By exposing the vehicle to a plethora of environments, including simulation, closed courses, and public roads, the safety demonstration stage will provide robustness testing for the entire system.

[a] Koopman and Wagner, 2016a.

NOTE: Figure compiled by Jeremiah Robertson.

The same behavioral competencies or scenarios outlined here and in the previous section ("A Discussion of the Setting Frame") could be conducted on a public road. Doing so would have the benefit of great validity. It also would allow for measuring how well an AV system can operate and be robust in the chaos of the public road—although such testing is not repeatable and will not occur with sufficient frequency to collect sufficient data for statistically meaningful inferences.[85] Because demonstration on public roads is occurring in a setting involving the general public, said public is now at risk.

Decisions about how to formalize and promote demonstration for purposes of measuring and communicating about safety imply some kind of governance—within the company in charge of AV development, through an industry body, or possibly by a government entity or third party. To date, the U.S. government approach to safety has involved self-certification by manufacturers that they comply with FMVSS (as outlined in Chapter 2). Were a demonstration protocol for AVs to be put forth by a governing body, it could require a method containing the following:

- an ensemble of scenarios, with vetting for relevance to key driving tasks
- adaptation of the protocol to the vehicle's ODD
- standardization around a specific protocol for running demonstration activities, understanding the vehicle's actions, and passing or failing
- certification power as resting in the developer, a government entity, or a third party
- the AV's ability to identify that it has exceeded its ODD and to devolve to minimal-risk conditions.

A protocol would likely draw from simulation, closed-course testing, and public-road testing, within increasing levels of validity and risk and decreasing levels of abstractness in each setting respectively.[86] It would have to be designed to minimize gaming of the protocol by developers, who can be expected to design for satisfying the protocol—which is different from addressing real-world requirements. Demonstration does not prove the vehicle is safe; at best, it shows the vehicle is not unsafe.[87]

Deployment

Deployment entails ongoing safety challenges. Maintaining the proper functioning of the software, hardware, sensors, and cameras becomes crucial. Software updates will occur regularly, as discussed in the "Upgrades" section later in this chapter. Hardware also might require occasional updates—old hardware is not always capable of running new software at peak effectiveness. Cleaning and aligning sensors and cameras present challenges, as discussed in the section "AV Safety and Business Model." Depending on the life span of the vehicle, all of these

[85] Kalra and Paddock, 2016.

[86] Koopman and Wagner, 2018.

[87] Nowakowski, Shladover, and Chan, 2016; University of California PATH Program, 2016.

might require replacement—attention to when a sensor is not working properly (or at all) will be even more important with AVs than it has been for conventional vehicles.

While crashes are probable, if expected to be uncommon, in the development and demonstration stages, the incidence during deployment will likely increase with greater exposure. Post-crash survivability for all road users depends on fast crash notification, fast arrival of emergency services, and fast transport to medical facilities. Around crash notification, advanced automatic crash notification (currently available in conventional vehicles) can improve survivability by notifying emergency services of a crash and providing severity details so that appropriate services can be dispatched.[88] Emergency medical services personnel might need special training around vehicle extrication techniques for AVs.

Once AVs are commercially available, mileage will accumulate with use. Increases in exposure generally improve the ability to measure safety, as discussed later in this chapter in the sections on "Frame 3. Measures," "From Frames to Framework," and "Measures to Metrics." Monitoring is especially important during early-stage deployment because people not associated with the development or demonstration process are likely to subject AVs to unexpected challenges and environments. People inside and outside the vehicle will challenge technology in unanticipated ways, sometimes creating safety hazards. Beyond this, AVs will fail in unanticipated ways. Conventional vehicle behavior might provide little guidance in anticipating these failures. Reconstructing the event and the AV's decisionmaking pathways will be critical to post-event investigation.[89] Event reconstructions provide valuable insight as case studies.

A Discussion of the Stage Frame

Stages occur in different settings; development and demonstration take place in artificial settings and on public roads, deployment only on public roads. Safety in development, demonstration, and deployment can be assessed by multiple measures, with different interpretations and different concerns. Each measurement can be obtained in each setting. However, the risk to which the public is exposed in order to obtain the measure and how that risk is controlled can vary at each stage.

Frame 3. Measures

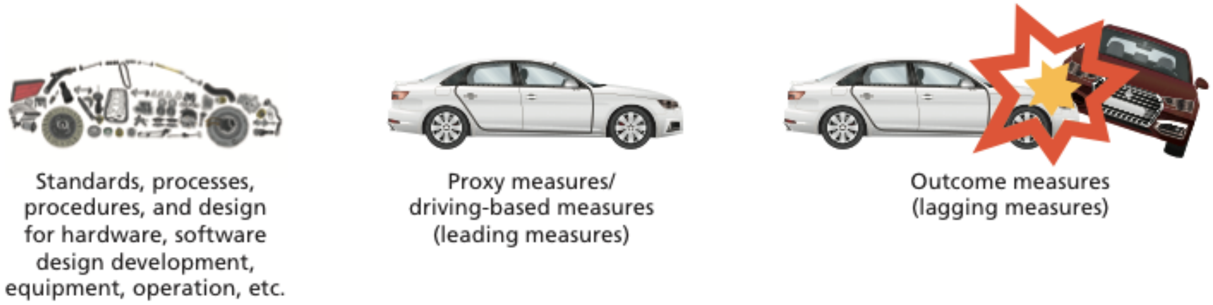
Three categories of measures exist around AV safety. These categories are defined by when, in vehicle construction and operation, they are available (Figure 3.4). *Measure* refers to concepts (e.g., crashes with severe injury) and are less precise and implementable in their definition than

[88] Ecola et al., 2018.

[89] Event data recorders, which capture information in the seconds before, during, and after a crash, are among the tools that can help, although they would be expected to capture more and different kinds of data with AVs than they do today. NHTSA, undated-c.

executable metrics (e.g., crashes with severe injury [rated at 3 or higher on the Abbreviated Injury Scale[90]] per 100,000 vehicle miles traveled). Metrics will be discussed later in this chapter, in the section "Measures to Metrics."

Figure 3.4. Measure Frame


Standards, processes, procedures, and design for hardware, software design development, equipment, operation, etc.
Proxy measures/ driving-based measures (leading measures)
Outcome measures (lagging measures)

SOURCE: Dezay, Yuri Schmidt/Adobe Stock.

Measure Category 1: Standards, Processes, Procedures, and Design

The first category consists of the standards, processes, procedures, and design requirements involved in creating the AV system hardware, software, and vehicle components. Processes and standards could be technology-specific or, to be inclusive, err on the side of being overly broad. These come from many sources, including FMVSS, ISO, and SAE, each with (slowly) evolving application to AVs. Safety driver training could be included here. Because standards, processes, procedures, and design are evolving constantly and might be measured at lower levels of vehicle composition (e.g., at the subcomponent level)—and because measures of adherence relate indirectly to safety through function—the focus falls to the second and third categories, leading and lagging measures, as will be discussed. That said, this category of measure shows the approaches through which it is possible to track how safety is built in, beginning with design. Adherence to industry standards is both industry best practice and associated with reducing potential liability for companies.

Measure Category 2: Leading Measures

Leading measures reflect performance, activity, and prevention; lagging measures are observations of safety outcomes or harm.[91] Leading measures draw from the pre-crash row of the Haddon Matrix (Table 2.1). These measures serve as proxies or surrogates for lagging measures. Leading behaviors associated with a safety outcome are not themselves an outcome, but in

[90] The Abbreviated Injury Scale is maintained by the Association for the Advancement of Automotive Medicine (undated).

[91] Government of Alberta, 2015; Occupational Safety and Health Administration, 2005; Occupational Safety and Health Administration, 2016.

changing the precursory behavior, the probability or severity of outcomes are affected. The link between a proxy and an outcome should not relate to the circumstances under which either occurred (e.g., a strong proxy when speed is high, a weak proxy when speed is low). Additionally, the link relates to risk of collision, not of injury, which would include other considerations such as fragility of the road user.[92] Because the events included in leading measures happen with greater frequency than events in lagging measures, leading measures are often obtainable with statistical confidence at lower cumulative mileage than lagging measures, giving them a canary-in-the-coal-mine quality.

Cumulative mileage is a possible leading measure, but there is no relation to safety without more information around the parameters and density of challenges. Also, system upgrades could mean that past performance might not reflect a system's current safety level.

Disengagements are another leading measure. A *disengagement* is the

> deactivation of the autonomous mode when a failure of the autonomous technology is detected or when the safe operation of the vehicle requires that the autonomous vehicle test driver disengage the autonomous mode and take immediate manual control of the vehicle.[93]

Disengagements can be initiated by the AV itself, by in-vehicle or remote safety drivers, or by passengers.[94] Disengagements are currently used as a nonstandardized safety measure by various companies, states (required of entities testing AVs in California), and even by the federal government.[95]

As a leading measure, disengagement's main benefit is that it is well established, although not uniformly defined, applied, or reported.[96] Diminishing disengagement frequency, although reflective of how well the system is learning, does not mean that the system is safe. Disengagement is sensitive to driver training, which might reflect company risk aversion, and to driver risk tolerance. It requires a safety driver and thus is not approach-neutral—for example, disengagements are not used in Tesla's shadow mode.[97] Using disengagement as a safety measure can produce perverse incentives; the safety driver might try to disengage as rarely as possible. Additionally, as the AV system improves, safety drivers could lose skill and become more likely to be distracted because they have decreasing need to act.[98]

[92] Johnsson, Laureshyn, and de Ceunynck, 2018.

[93] California Code of Regulations, undated.

[94] California Department of Motor Vehicles, undated.

[95] California Department of Motor Vehicles, 2017.

[96] Felton, 2018.

[97] Prescott, 2017.

[98] That reality, echoing situations observed with aviation automation, has motivated the rise of driver monitoring systems.

Simulating a situation post-disengagement to explore what could have happened had the disengagement not occurred could facilitate understanding about whether the disengagement was warranted.[99] The simulation might suggest whether a crash would otherwise have occurred. But such post hoc simulation shifts the variability inherent in human test drivers on public roads to variability based on simulation quality and ability to predict a counterfactual. Additionally, if safety drivers believe that the only disengagements that will count are the ones that simulations show were needed, those drivers might be overly liberal in disengaging.

Infractions are another leading measure.[100] Infractions could be recorded by the vehicle or police. The former is preferred because the latter is heavily mediated by enforcement, which depends on whether the questionable activity is noticed and triggers a response. A statistically significant relationship exists between infractions and later crashes for human drivers, and it is assumed at this early time in AV history that such a relationship could apply to AVs.[101] Right-of-way infractions, including "failure to yield" and disobeying traffic signals, have the strongest relationship to later crashes.[102] Infractions as a measure might benefit from removing infractions involving judgment (e.g., reckless driving), infractions that are not germane to AVs (e.g., distracted driving), infractions unrelated to safety (e.g., parking infractions), or state- or county-specific infractions. However, circumstances could occur in which an action that results in an infraction is the safest option, such as speeding or crossing unbroken white lines to avoid a crash. This measure is unable to distinguish between infractions that, although technically illegal, enhance safety and infractions that worsen safety. Additionally, the relationship between infractions and crashes for human drivers might not apply to automated systems.

Roadmanship captures the ability to drive on the road safely without creating hazards and responding well (regardless of legality) to the hazards created by others.[103] The concept centers on whether the vehicle "plays well with others," even if others are not around.[104] Roadmanship is an attractive leading measure of safety conceptually but complicated definitionally. A roadmanship measure should be

- objective
- physics-based
- available using current technology

[99] Waymo factors such post hoc simulations into its disengagement reports to the California Department of Motor Vehicles. Waymo, 2016.

[100] Infractions could also include violations or misdemeanors. Exact terminology differs between states. Within this report, *infractions* refers to noncriminal violations of state and local traffic law.

[101] Chen, Cooper, and Pinili, 1995; Lightstone, Peek-Asa, and Kraus, 1997; Lui and Marchbanks, 1990.

[102] Chen, Cooper, and Pinili, 1995.

[103] The concept is not formally defined because of a previous lack of measurement capabilities and lack of need with conventional vehicles, for which crashes are plentiful.

[104] Koopman and Wagner, 2018; Sagar Behere, senior manager, systems engineering—autonomous driving, Toyota Research Institute, comments at RAND workshop, May 15, 2018, and follow-up discussion with authors, August 19, 2018.

- reflective of the official and unofficial rules of the road.

Additionally, a roadmanship measure should separate the initiator of an unsafe decision from a responder (so as not to punish evasive action) and should reward predictability and anticipatability. This measure should capture any chaos around the AV, and ideally whether that chaos can be attributed to the behavior of the AV or that of another road user. Table 3.2 presents a noncomprehensive list of established measures with roadmanship features. Variations exist on each of the examples listed. Measures around only specific road users or specific crash scenarios were not included.

Of all the examples listed, the safety envelope is closest to the concept of roadmanship.[105] But how can envelopes based on maneuverability, infrastructure, velocity, time, or space be determined? The safety behavior rules comprising the Responsibility Sensitive Safety (RSS) model is one example of how the envelopes could be defined.[106] This model provides algorithmic proofs around how to, in a variety of environments, (1) maintain a safe distance in front of the vehicle, (2) maintain a safe distance lateral to the vehicle, (3) respect and give right-of-way, and (4) exercise caution when a sensor's perception is occluded because of infrastructure or road design. The first two rules feed into envelopes based on velocity, time, and space; the third feeds into maneuver-based envelopes, and the fourth into envelopes based on infrastructure and roadway environment (although this could be expanded to other infrastructure if needed). In this context, safety envelope violation counts signify the AV's ability to follow rules of the road. To fully meet the definition of safety envelopes (and hence roadmanship), the RSS model includes the calculable concept of responsibility, allowing discernment between initiators of a safety envelope violation and responders. To implement this concept, a method of tracking violations of the safety envelope is needed.

The number of pathways to an incident is infinite,[107] but one of a finite number of driving behavior rules is almost always violated. Somehow, safety envelopes were violated—one of the RSS rules was not met by the AV or another road user. There could be other attractive ways to define the safety envelopes and other attractive measures through which to consider roadmanship overall.

[105] The concept has been used in other domains, from aviation to nuclear power.

[106] Shalev-Shwartz, Shammah, and Shashua, 2017.

[107] This concept reflects the *Swiss Cheese theory* of crash causation: Many individual events contribute to the likelihood of a crash, and when all the "holes" align, a crash occurs. See Ecola et al., 2018; and Perrow, 1999.

Table 3.2. Examples of Measures Around Roadmanship

Category	Detail
Near misses or near crash	
Definition	A subjective judgment that a circumstance had the potential for a collision
Example	Hayward, 1972; NHTSA, 2006; Klauer et al., 2014; Uchida et al., 2010; Arai et al., 2001
Strengths	Frequently used
Weaknesses	Subjective; not uniformly implemented
Rapid acceleration or deceleration	
Definition	A change in velocity at or over a given quantity (These measures can be coupled with *jerk profile*, the rate of change of acceleration, and *yaw rate*, the rate of change of the angular velocity, and are related to delta V, or change in velocity.)
Example	Arai et al., 2001; Stipancic, Miranda-Moreno, and Saunier, 2018; Johnsson, Laureshyn, and de Ceunynck, 2018; Mahmud et al., 2017
Strengths	Objective; physics-based; available using current technology
Weaknesses	Fails to distinguish initiators from responders; incomplete; alone it could fail to fully account for lateral movement; relationship to lagging measures dependent on driver abilities and brakes.
Time to collision	
Definition	The length of time until a collision, should the vehicle continue in its current path (One could also sum the number of occurrence when the time to collision fell below a given number of seconds. Similar measures exist around distance between two road users instead of time.)
Description	Johnsson, Laureshyn, and de Ceunynck, 2018; Mahmud et al., 2017; Federal Highway Administration, 2008
Strengths	Objective; physics-based; available using current technology; currently in use in simulation and public road settings
Weaknesses	Fails to distinguish initiators from responders; incomplete; might not serve as a proxy for all crash types; relationship to lagging measures dependent on driver abilities and brakes.
Post-encroachment time	
Definition	The time between one road user departing from a location of potential collision to the time another road user arrives in that same area
Description	Nadimia, Behbahania, and Shahbazib, 2016; Mahmud et al., 2017; Federal Highway Administration, 2008
Strengths	Objective; physics-based; available using current technology; currently in use in simulation and public road settings
Weaknesses	Fails to distinguish initiators from responders; incomplete; might not serve as a proxy for all crash types; relationship to crash dependent on driver abilities and brakes.
Instantaneous safety metrics	
Definition	The probability of an unavoidable crash, recalculated at every instant (This measure is conceptually similar to the Crash Propensity Metric [Wang and Stamatiadis, 2014], the probability of a conflict devolving into a crash, calculated from simulated conflicts, and human and vehicle variability.)
Source	Every et al., 2017
Strengths	Objective and physics-based
Weaknesses	Fails to distinguish initiators from responders; currently still in development

Category	Detail
Combined indicators	
Definition	A generic definition of counting safety-critical incidents where, should each road user's course remain unchanged, a crash would occur (There are several ways to conglomerate conflict. One approach, known as Conflict Severity, combines maximum average deceleration and time to collision, both reflecting the effectiveness of evasive maneuvers, and delta V, or change in velocity.)
Description	Johnsson, Laureshyn, and de Ceunynck, 2018
Strengths	Generally available using current technology, although some methods might use techniques not yet available
Weaknesses	No universal, accepted, and vetted measurement-generation method; weaknesses vary based on method; some are subjective, some are incomplete, etc.
Safety envelope violation	
Definition	Counts how often an AV's safety boundary is violated, who is at fault, and how quickly the boundary is restored (Around the AV is a boundary defined by a measure of space, velocity, time given a speed to collision, etc., between the AV and other roadway users or objects. The AV might have a series of boundaries around it, wherein violation of each successive boundary provokes a more extreme evasive response by the AV. Envelopes can also be defined by infrastructure—e.g., lane markings—or by maneuvers—e.g., disallows turns at speeds where traction will be lost—limiting the vehicle in what it can do.)
Source	This concept comes from aviation (Lombaerts et al., 2015), but has been included in AV research (Koopman and Wagner, 2018)
Strengths	Meets aforementioned characteristics
Weaknesses	Agreed-upon definitions of safety envelopes do not yet exist in the automotive industry

Measure Category 3: Lagging Measure

Lagging measures involve actual harm—crashes—and their outcomes. This category draws from the second (crash) and third (post-crash) rows of the Haddon Matrix (Table 2.1). Severity-based outcomes include any contact between the vehicle and its outside environment (e.g., other road users, animals, or property), crashes resulting in property damage over a certain cost, crashes resulting in injury, in severe injury (Abbreviated Injury Scale 3+), in death, etc. As an additional layer, outcomes can be classified by crash types or configurations. Measures that integrate across crash severities include cost (e.g., medical expenses, missed work, and property repair), disability-adjusted life years, and quality-adjusted life years.

Measure Characteristics

There are many possible characteristics that a measure of AV safety should contain; most can be divided into four categories: valid, reliable, feasible, and non-manipulatable (Table 3.3).[108]

[108] We solicited inputs from people we interviewed (from the public and private sectors) and generated a much longer set of characteristics. We used our judgment to compress the list to these four, recognizing that too many characteristics would not be practical to use.

Valid, specifically in the sense of *construct validity* (a type of internal validity),[109] refers to the extent to which the evidence correlates with safety. *Reliable* refers to how consistently, quantitively, and objectively measurements be made, and how stable those measurements are. *Feasible* refers to whether measurements can be obtained with reasonable time, cost, etc. *Non-manipulatable* refers to being difficult to tamper with or game.[110]

Table 3.3. Measuring Characteristics

Characteristic	High	Medium	Low
Valid	Metric directly measures characteristic	Metric is closely related to characteristic	Metric is indirectly related to characteristic
Reliable (between AV models and within AV models; comparing vehicle to vehicle)	Quantitative, objective, well-defined, and stable	Somewhat quantitative, subjective, well defined, or volatile	Qualitative, subjective, volatile, and anecdotal
Feasible	Evidence can be gathered with reasonable time, cost, and resources	Evidence could be collected with some difficulty	Evidence would be challenging to collect
Non-manipulatable	Provides no opportunity for manipulation of measure	Some opportunity exists for manipulation	Easily manipulated

A Discussion of the Measure Frame

Each measurement can be obtained in each setting. However, the risk associated with obtaining the measurement can vary, as can the quality of the measurement (e.g., the validity, feasibility, reliability). For example, a developer can obtain an outcome measure, such as a crash rate from 1,000 miles in a simulated urban environment and a crash rate from 1,000 miles in downtown San Francisco, California. In the former, only simulated people, animals, and property are at risk. In the latter, real people, animals, and property are at risk. Additionally, depending on the simulator, the simulated rate might have little relation to the rate obtained from driving on public roads with a safety driver, meaning the simulated rate had low validity. However, if the simulation were run 5,000 times, always with the same result, the simulated rate could have high reliability. Similarly, safety in the development, demonstration, and deployment stages can be measured using leading and lagging measures, but a given measurement's meaning and quality (e.g., validity, feasibility, etc.) could differ at each stage.

[109] Hartas, 2015.

[110] Savitz et al., 2015.

From Frames to Framework

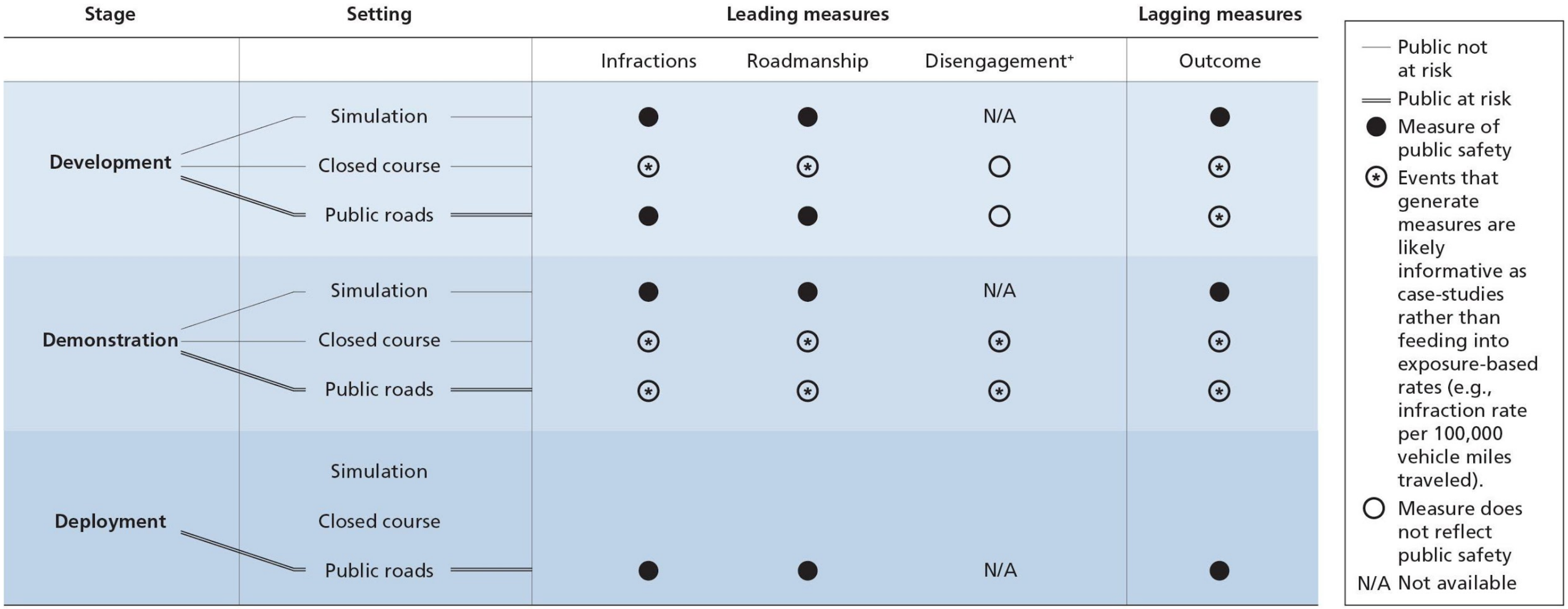
Settings, stages, and measures are integrated into one framework in Figure 3.5. The table lists lagging and leading measures at combined stages (development, demonstration, deployment), settings (simulation, closed course, public roads), and public risk. The framework shows what can be measured and how each measure can be used to make a case for the safety of the AV at each stage and in each setting.[111] A black circle or circle around an asterisk indicates measures that reflect public safety in a way that could be communicated publicly. An open circle indicates measures that do not reflect public safety but could be used internally by a company to measure system function and progression. An N/A (not applicable) indicates that measures cannot be calculated in this setting within this stage.

Focusing on the rows in Figure 3.5 (because the emphasis is on the safe conduct of these vehicles for the public), the first two rows are not themselves germane measures of public safety because the public is never exposed to risk—rather, these rows reflect how well the vehicle functions. The same argument could be made for the fourth and fifth rows because, again, the public is not at risk. But because these rows serve as formal, semiformal, or informal gateways for development on public roads or deployment, they are determinative of safety, even if they are not themselves situations in which the public is actually exposed.

Focusing on the columns, it is possible to measure safety in most cells. Considering disengagement, in simulation there is no disengagement, so this measure is not applicable within the development stage in this setting. During the development stage, the public and regulators should focus on whether the development process is safe for the public, not whether the AV system is progressing or learning. This perspective flags disengagement as a poor measure of safety during development on closed courses and public roads. As discussed in the earlier section "Measure Category 2: Leading Measures," disengagement is misleading without context—there are perverse incentives, and it is subject to the safety driver's training and risk tolerance. Overall, the public does not need to consider why the crash did not occur, just that it did not. The "why"—either because of an excellent AV system performance or disengagement—is the developer's purview. In the demonstration stage on closed courses and public roads, disengagement becomes informative if a human safety driver is involved (compared with an unmanned vehicle) because it suggests the AV might not be ready for commercial deployment. Additionally, because a safety driver might disengage and avoid a crash, infraction, or error in roadmanship, disengagements must be considered in the demonstration stage to allow a complete

[111] NHTSA, 2017b; NHTSA, 2017c.

Figure 3.5. Integrated Safety Framework



Stage	Setting	Leading measures			Lagging measures
		Infractions	Roadmanship	Disengagement[+]	Outcome
Development	Simulation	●	●	N/A	●
	Closed course	⊛	⊛	○	⊛
	Public roads	●	●	○	⊛
Demonstration	Simulation	●	●	N/A	●
	Closed course	⊛	⊛	⊛	⊛
	Public roads	⊛	⊛	⊛	⊛
Deployment	Simulation				
	Closed course				
	Public roads	●	●	N/A	●

—	Public not at risk
═	Public at risk
●	Measure of public safety
⊛	Events that generate measures are likely informative as case-studies rather than feeding into exposure-based rates (e.g., infraction rate per 100,000 vehicle miles traveled).
○	Measure does not reflect public safety
N/A	Not available

[+] This column assumed that, in the closed course and public road settings, a safety driver is available (either in the vehicle or remotely). If a safety driver is not present, this entire column would be N/A.

view of safety.[112] In deployment, in lieu of disengagement, the system should be able to transition to a minimal risk-condition when it suffers a technical failure or a departure from its ODD.[113] Failure to do so suggests deep problems in the perception or planning subsystem and would likely result in a lagging measure event (i.e., a crash).

Moving to other measures, safety outcomes, infractions, and roadmanship are germane in every stage and setting. However, use and meaning by stage and setting varies (Box 3.2). The chief determining factor is exposure. Those combinations of measure, stage, and/or setting denoted by a circle around an asterisk (rather than a black circle) in Figure 3.5 are interpretable as a case study (rather than as an exposure-based rate). In situations of low exposure, events can be leveraged as case studies. For example, exposure on the order of millions or billions of vehicle miles traveled would be needed to detect even modest differences with statistical confidence in crash rates between conventional vehicles and AVs on public roads.[114] It is extremely unlikely that within "development, public roads" there would be enough exposure to calculate such a difference within a reasonable amount of time. However, because vehicle-recorded infractions are likely to be more common than crashes, a lower and potentially attainable mileage is needed to obtain exposure-based measures, such as infractions per VMT, and to generate with statistical confidence comparisons with infraction rates among conventional vehicles during development on public roads. Unfortunately, such infraction rates do not currently exist for conventional vehicles, an issue covered later in this chapter in the section titled "Creating Comparisons."

The safety measure's interpretation is parameterized by where and when the measurement occurs. As the field of AV safety matures, this matrix can be expanded, contracted, and refined. Figure 3.5 is a living document.

[112] Today, other than case studies of lagging events of high severity, disengagements are the only safety measure publicly available. Within the report we separate development (when disengagements are not informative) from demonstration (when disengagements are highly informative and central to a comprehensive view of safety). The demonstration stage as outlined in the report does not currently exist. Until the demonstration stage is distinct, clearly defined, and in use, there might continue to be value in looking at disengagements, if well defined and reliably measured.

[113] A deployed Level 4 AV would not have the safety driver used in its development. It would be limited to a specified ODD, and either a way to excuse itself—retreat to a safe space—or invoke remote assistance if outside of its ODD would be expected.

[114] Kalra and Paddock, 2016.

Box 3.2. Interpretation of Each Measure at Each Stage in Each Setting and Potential Rating of Characteristics (outlined in Table 3.3)

The interpretation of each measure varies by stage and setting. Consider the lagging measure of all crashes with a minimum property damage of $5,000. Measure meaning varies:

- In development using simulation, the measure reflects the count or rate per given simulation (or per dangerous scenario encountered, etc.) of simulated crashes. Because no actual damage or injuries can be sustained, this value is hypothetical and results from the developer estimations or conjecture. Pursuant to simulation quality, this could suggest the vehicle's ability to handle scenarios in a controlled environment.
  - Such a measure might have medium validity, medium to low reliability, high feasibility, and low non-manipulatability.
- In development using a closed course, the measure reflects the count of crashes where the hypothetical damage to the course or vehicle was $5,000 or more, within specific parameters. This value is hypothetical. Course operators use balloon cars and dummies as target objects, so the hypothesized values would represent the probable cost had road and other road users been real. The cost of damage to the AV also must be hypothesized because the AVs should not suffer nearly as much damage (if any) as they would in public road crashes. This could indicate the vehicle's ability to handle scenarios in a semi-controlled environment.
  - Such a measure could have low validity, low reliability, medium feasibility, and medium non-manipulatability.
- In development on public roads, the measure reflects the detailed crash investigation results and count of crashes where damage to the AV, road users, or property cost $5,000 or more. This could indicate the vehicle's ability to handle the challenge of an environment within ODD specifications while minimizing the hazards put upon the public. Conclusions about the vehicle's ability to handle challenges are caveated by the presence of a safety driver, who could (or should) disengage the AV system, although that is not always sufficient for crash avoidance.
  - Such a measure might have high validity, medium reliability, high feasibility, and high non-manipulatability.
- In demonstration using simulations, the measure reflects the count or rate per given simulation (or per scenario) of simulated crashes, where severity of crash has also been simulated, according to a prespecified protocol. As with development actions in simulation, this value is hypothetical. This could indicate the vehicle's ability to handle the scenario in a controlled environment to a predetermined standard and the vehicle's readiness for public deployment.
  - Such a measure might have medium validity, medium reliability, medium feasibility, and medium non-manipulatability.
- In demonstration using a closed course, the measure reflects the count of crashes where damage to the course or vehicle was $5,000 or more, within predesignated maneuvers or scenarios. As with development actions on closed courses, this value is hypothetical. This could indicate the vehicle's readiness for public deployment.
  - Such a measure might have medium validity, medium reliability, medium feasibility, and medium non-manipulatability.
- In demonstration on public roads, the measure reflects the detailed crash investigation results and count of crashes where damage to the AV, other vehicles, other property, or other persons cost $5,000 or more. This could indicate the vehicle's ability to handle predesignated maneuvers or scenarios in an environment with little control exercised by developers or testers beyond ODD specifications. This might strongly indicate the readiness for public deployment.
  - Such a measure might have high validity, medium reliability, high feasibility, and high non-manipulatability.
- In deployment on public roads, the measure reflects the detailed crash investigation results and rate of crashes where damage to the AV, other vehicles, other property, or other persons cost $5,000 or more. This could indicate areas of improvement for future upgrades.
  - Such a measure might have high validity, medium reliability, high feasibility, and high non-manipulatability.

Measures to Metrics

No single measure at any stage or setting can tell the entire story of safety. As described in the earlier section "From Frames to Framework," each measure has its own interpretation and relationship to other measures. Leading measures serve as canaries for lagging measures, the rarer events. Lagging measures can also be canaries for other lagging measures. For example, a case study of minor contact between an AV and a scooter in the development stage might augur more-frequent and more-severe crashes during deployment. Combining measures might offset each one's individual strengths and weaknesses. This is especially true when measures (concepts) become metrics (a defined calculation). For example, juxtaposing one metric that has high non-manipulatability but medium validity and feasibility with another metric that has high validity and feasibility but medium non-manipulatability might present a coherent safety argument.

At every stage in each setting, there are many ways in which measures can become metrics (example metrics are presented in Table 3.4). A measure can take many forms: A crash can be a case study or included in a count or a rate; the cost can be added to total costs; or the resultant injuries can contribute to days in a hospital, days of work missed, disability-adjusted life years, quality-adjusted life years, etc. Similarly, an infraction can be a case study or included in a count or a rate with varying denominators such as number of scenarios in simulation, cost of a police ticket (had the police caught the infraction), etc. Metrics that do not have exposure (denominator) parameters do exist—as counts, case studies, cumulative costs, or binary results (e.g., pass or fail). But when making comparisons, exposure provides necessary context and narrative.

Conventional vehicles frequently form the safety benchmark for AVs—asking how much safer than conventional vehicles AVs should, could, or will be. It is natural to use the status quo of conventional vehicles as a benchmark.[115] Comparisons of AVs with conventional vehicles or with other AVs generally require event data (numerators) and exposure data (denominators) for both types of vehicles.

[115] Schoettle and Sivak, 2015; Teoh and Kidd, 2017; Blanco et al., 2016.

Table 3.4. Selected Measures (Numerators), Exposure (Denominators), and Comparisons for AV Safety Metrics and Potential Rating of Characteristics (outlined in Table 3.3)

Measure	Exposure	ODD	Comparisons	Valid	Reliable	Feasible	Non-Manipulatable
Deaths (within 30 days of crash)	Per VMT (on public roads)	Non-highway roadways and parking lots in urban Dallas, Texas, during daylight hours, with no precipitation, at all, in all traffic conditions	All conventional vehicles and human drivers in urban areas from 7 a.m. to 7 p.m. in normal weather[a]	High	High	Medium[b]	High
Serious injuries (Abbreviated Injury Scale 3+)	Per VMT (on public roads)	Highways in California and Oregon in all weather conditions and all speeds under low traffic conditions	Specific subgroup of human drivers and conventional vehicles[c]	High[d]	Medium	Medium[b]	High
Serious injuries (Abbreviated Injury Scale 3+) (simulated)	Per VMT (simulated)	All roadway types in clear weather or rain at speeds up to 35 mph	Specific subgroup of human drivers and conventional vehicles in clear or rainy weather on roadways with a speed limit at or below 35 mph[c]	Medium	Low[e]	Low[e]	Low[e]
Crashes (Collision Deformation Classification code [SAE J224] of 2 in terms of maximum extent of penetration)	Per hour driving (on public roads)	All roadway types, in all weather, at all speeds, with clear lane markings and no pedestrians, bicyclists, or scooters detected	Same model–year conventional vehicles with ADAS, excluding bicyclist and pedestrian crashes[c]	Medium[d]	Medium[f]	High	High
Safety envelope violations (reflecting roadmanship)	Per licensed driver or customer	Fourteen city blocks in downtown Seattle, Washington, on Fifth Street from Olive Way to Terrace Street	Other AV makes and models in the same location[g]	High	Medium[f]	High	Medium[f]
Disengagement	Per scenario on closed course	Suburban environment in clear weather, snow, or rain, with visible or disfigured signage (as created in a closed course)	Pre-upgrade AV of same make and model of the same closed course, three years ago[g]	Low	Medium	High	Low
Vehicle-counted safety-related infractions	Per urban city block (on public roads)	Washington, D.C., in all weather conditions, in areas with streetlights, excluding roundabouts	Prespecified standard or number[g]	Medium	Low[h]	Medium	Medium

[a] Fatality Analysis Reporting System (FARS).

[b] FARS and National Automotive Sampling System/Crashworthiness Data System (NASS/CDS) data are not timely.

[c] NASS/CDS.

[d] The metric's ODD is not the same as or similar to the comparison group's, inviting bias.

[e] Injuries can be very difficult to simulate accurately.

[f] In both these measures, the denominator is highly susceptible to confounding.

[g] Data source does not currently exist.

[h] Without more information or standardization on closed-course loops, this measure is not reliable.

Numerators and Denominators

Numerators and denominators pull from the measures outlined in Table 3.4 and beyond. The most popularly touted denominator is VMT, simulated and on public roads. However, as already discussed in the section "Measure Category 2: Leading Measures," not all miles are the same. Details on the VMT context and content are needed to control for confounding.

The challenge for numerators and denominators lies in data from conventional vehicles. Such data often lag by a year or more in availability, and, around the numerator, might be limited to high-severity incidents (e.g., police-reported crashes are generally limited to those involving an injury or a given amount of property damage). Additionally, because AVs' ODDs can vary, comparable conventional vehicle data must precisely map to the same parameters. Parameters can include weather, time of day, road environment, etc. Business models also must be considered—for example, driving patterns for ride-share programs could differ from driving patterns for private vehicles. Obtaining such precisely sliced data for conventional vehicles is challenging.

Creating Comparisons

Comparisons (Table 3.4) should always contrast like with like. Otherwise, the comparison is biased. This concern gives weight to the question of which conventional vehicles should be included. Contrasting an HAV to a 12-year-old car (the current average vehicle age on the road in 2016[116]) is biased because of the advances that automotive technology has made in crashworthiness and post-crash survivability alone. Should AVs instead be compared with the newest conventional vehicles with the latest crash avoidance technology?[117]

Broadly, there are three categories of metrics when comparing AVs with conventional vehicles: can compare, could compare, and can't compare. *Can compare* consists of metrics for which comparisons can be made using available data. Police-reported crashes per VMT and police-reported infractions fall into this category, assuming conventional vehicle exposure data are available specific to the ODD. Besides VMT, metrics could be per licensed driver or per passenger group (when there is no driver). Where traffic cameras are available to capture specific scenarios, per-scenario-based rates can be computed. This category is the most likely to inform equivalent risk requirements for AVs compared with conventional vehicles.

Could compare consists of metrics where comparisons are possible, but data are not currently available. These include all measures from simulation and closed courses (because conventional vehicle measures are unlikely to be already available in these settings) and crashes that fall below the police reporting threshold for severity (because such data are not available for

[116] Statista, 2017.

[117] Kalra and Paddock, 2016.

conventional vehicles). Most proxy measures fall in the *could compare* category: Conventional vehicles do not currently collect data on roadmanship or vehicle-recorded infractions. Naturalistic driving data, obtained by instrumenting vehicles to unobtrusively monitor the driving experience inside and outside the vehicle in its "natural setting,"[118] could fill this need. Such naturalistic data must be available from a similar ODD and from a representative sample of drivers. Specific groups of drivers (e.g., commercial drivers or those in safety-promotion programs) might be safer than the average human driver, creating a conservative sampling bias.

Can't compare consists of metrics for which comparisons to conventional vehicles are not possible. The most obvious measure is disengagements, which do not occur with conventional vehicles. Arguably, because each developer reports (and possibly records) disengagements differently, this metric also cannot be used to compare different AVs.

Operational Design Domain

ODD refers to the conditions in which the developer intends the AV system to operate; the "where" and the "when."[119] As referenced in Chapter 2, ODD is particularly important for Level 4 vehicles because the system is expected to be able to operate (monitor the driving environment; steer; accelerate or decelerate; and respond effectively where it cannot do those things— i.e., fall back) without the aid of a human driver within its ODD.

Developers need as thorough as possible an understanding of what the AV can and cannot handle—ideally determined in the artificial setting before the AV starts on public roads. A company could refine the ODD during development on public roads and forward through deployment. Because a Level 4 AV is responsible for keeping itself within its ODD, the ODD's precise definition must be in terms that are identifiable or inferable by the AV system.

Precision in ODD definition is complicated. Theoretically, the definition should involve inclusions and exclusions. But guaranteeing coverage in areas included or not excluded might be unachievable because a developer cannot test every possible scenario permutation (as discussed in the section "A Discussion of the Setting Frame"). Definitions should specify vehicle maneuvers and environment, but some maneuvers might be unachievable in certain environments. A maneuver-based ODD specifies what an AV can and cannot do (e.g., it can turn right at an intersection but cannot turn left). For an environment-based ODD, a distinction can theoretically be made between static and dynamic ODD environments. Dynamic components are subject to change: weather, traffic, etc. Static components are unchanging: roadways, infrastructure, neighborhoods, etc. But static ODD environmental components can change because of road work, police-directed intervention, or infrastructure decay. For AVs to drive on

[118] Uchida et al., 2010.

[119] SAE Mobilus, 2018.

public roadways, the public and policymakers must accept a level of uncertainty around AV performance even within an ODD.

ODD definitions describe minimal-risk conditions, which should occur when the AV finds itself outside of its ODD. Attaining minimal risk conditions involves having an in-vehicle occupant or remote driver take over the driving task, proceeding slowly, or stopping in a safe place.[120] As stated in the earlier section on "A Discussion of the Setting Frame," devolving to some conditions might present a new safety threat.

Common terminology for ODDs would facilitate intra- and inter-organizational communications, which would decrease the likelihood of surprises in terms of ODD coverage. A taxonomy of ODD attributes would support stakeholder communication about AVs. At a minimum, such a taxonomy could contain considerations of weather, other road users likely to be present, road condition and markings, time of day, roadway surroundings, traffic density, and AV system familiarity with the roadway. The level of taxonomic detail could vary depending on what consumers need to understand, what regulators need to know, and what information companies need to communicate internally—or, possibly, externally within the industry. Although taxonomic guidelines found in SAE J3016 map levels of automation to ODD,[121] bases for describing ODD are included neither within that recommended practice nor by NHTSA.[122] Developers might have constructed internal taxonomies, but a shared language would facilitate communication with regulators and consumers not just about what the AV can do but also when and where it can do it. Language must advance along with (or even in front of) technology, easing its path forward.

As described in the section "Numerators and Denominators," safety metrics include consideration of ODD in the numerator, denominator, and comparison. Caution is needed because comparing vehicles operating in different ODDs invites potential bias. If the comparison includes ODDs containing different base levels of safety threats, the vehicle in the safer environment is favored. For example, three miles of rural road at midnight have a crash risk profile that is different from three miles of road in a dense urban center during rush hour.

Upgrades and Updates

Once an automated driving system is in use, developers can use information gleaned, coupled with internal innovations, to change or upgrade the system. This can occur in any stage but is of particular concern in deployment because that stage continues in perpetuity; it is not bookended by another stage. Updates, notably to AV software and supporting services (such as mapping), are likely to occur frequently.

[120] SAE Mobilus, 2018.

[121] SAE Mobilus, 2018.

[122] NHTSA, 2017b.

Upgrades and updates present a large challenge to measuring AV safety using exposure-based measures. Within these measures, time is cumulative. For example, a measure of infractions per VMT reflects the infractions and VMT gathered over a formally defined time period. Lagging and leading measures are longitudinal (vehicle performance is followed over time) rather than cross-sectional (reflective of a specific moment). In longitudinal data assessment, the AV system must be stable. However, updates and upgrades disrupt stability. A tension exists between safety metrics reflecting the most recent safety level and safety metrics needing to draw from a period where the AV system is steady.

Longitudinal data analysis methods exist to model statistical comparisons of safety over time, even as safety levels change. However, these methods generally require defined, regular time periods (e.g., every two months) and for the AV system to be roughly equivalent within a time period. This is not realistic. Such methods are also not intended to track long-term safety trends. Alternative statistical modeling methods allow for differing lengths of time to be analyzed, but still require time to be bucketed. Bucketing exposure requires developers to decide when the system has changed sufficiently to start measuring safety afresh (i.e., to start a new bucket).

One possibility is to start accumulating events and exposure afresh every time an upgrade occurs. But with upgrades to AV software and supporting systems (e.g., maps) potentially occurring even daily, sufficient exposure would never accumulate to permit exposure-based metrics. An alternative would focus only on major upgrades, but that raises the question of what defines a major upgrade. Using lines of code can be misleading because code length does not translate to system complexity or change. Additionally, lines of code are less germane when using machine learning. The developer could declare when a major upgrade occurs, but such a method is heavily prone to manipulation and marketing pressure, and it has no guarantee of reliability or consistency.

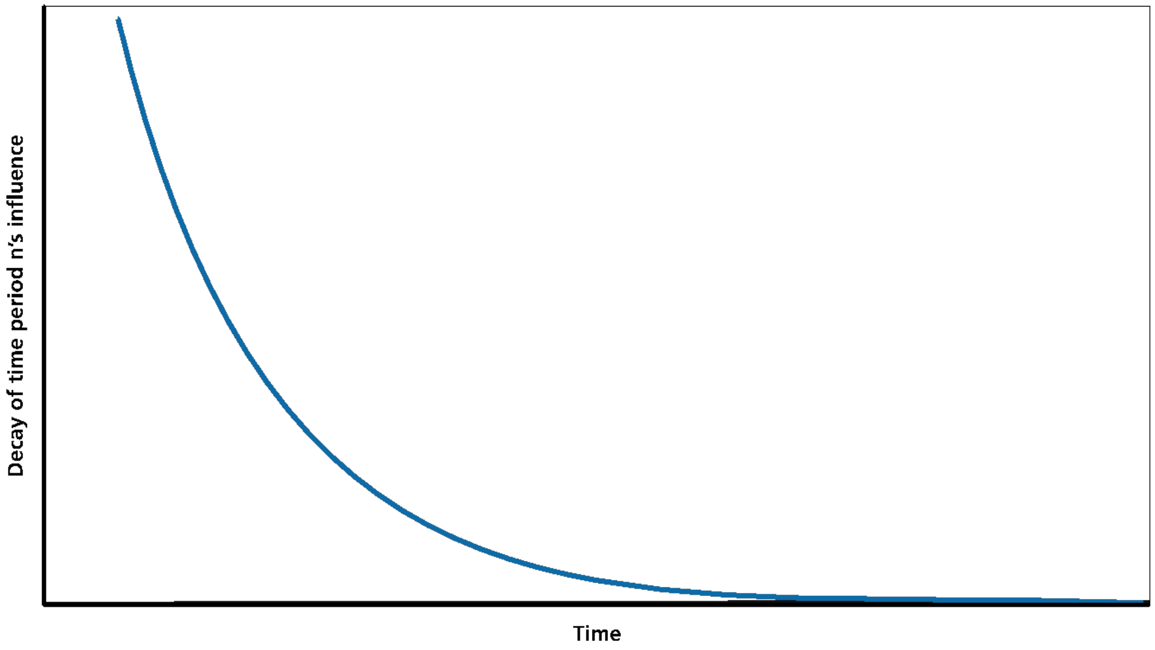
An alternative approach is to use a moving average over time instead of bucketing exposure. Using this approach, AV safety metrics always reflect the recent past (e.g., two weeks, three months, two years).

Metrics of safety in an upgradable environment should forgive but not forget. To have more forgiveness but less forgetfulness, an exponential moving average could be used, with the most recent period of time heavily weighted. This recurrence relation can be considered a form of backward discounting. For example, safety metrics from the past month could be weighted 60 percent, with the safety measures from previous cumulative time weighted 40 percent. The larger the weighting for the most recent period compared with the weighting of previous periods, the more forgiving the metric becomes. However, it can never truly forget because the past is always factored into in the calculation. This relationship can be expressed as:

$$A_n = W * D_n + (1 - W)A_{n-1}$$

in which the current average metric (A_n) is a function of the weighting (W) times the safety metric from the most recent period (D_n) plus the complement of the weighting ($1–W$) times the previous average of $A_{n–1}$, and n indicates the iteration. This formula allows all previous readings to remain in the average, subject to exponential decay in influence over time (Figure 3.6). Measures resulting from this formula could be used internally by the developer to track safety changes over time, to make their case for safety during demonstration, and as a reportable, ongoing measure during deployment. Such a measure could reflect the system becoming safer or less safe over time without bias. These moving-average approaches also are neutral to when the upgrade actually enters the system. Upgrades could be downloaded immediately if they occur over the air, but if the upgrades require a service call to a dealer, there will be larger delays and larger variations in when the upgrade is downloaded.

Figure 3.6. Decay of Influence of Time Period *n* over Time



The Ecosystem

Thus far, safety has been considered at the system level. As Figure 1.2 shows, the ecosystem level sits one step above the system level in granularity. Safety at the ecosystem level refers to the comprehensive burden of injuries resulting from road traffic crashes. Measuring safety at the ecosystem level uses an ecological approach.[123] Within the ecological approach, a community or geographic area, rather than an individual vehicle, makes up the unit of analysis. The safety

[123] Stern and Fineberg, 1996; Morgenstern and Thomas, 1993; National Institutes of Health, 2005.

metrics of two or more areas are compared, one or more of which contains AVs.[124] Ecological studies could reflect the safety benefit of communication between AVs and infrastructure or other vehicles (V2X); the ecological approach is expected to grow in complexity and spread in usage because they capture the safety behavior of multiple types of roadway users.[125] In the ecological approach, the vehicle-specific exposure measure is not considered directly, although more exposure in terms of mileage or higher AV penetration could result in a stronger association with safety.

Because ecological studies look at safety on the community level, leading measures (such as police-reported infractions) or lagging measures (such as police-reported crashes) can be used. Measures that require additional effort to collect or are available only on some vehicles (e.g., low-severity crashes or roadmanship) might be complicated or impossible to use. If multiple developers release their AVs onto public roads in the same geographic area, this method will not allow analysts to discern the impact from each developer because data are measured on the community level rather than the vehicle level.[126]

AVs can be considered to operate on three levels of safety. The first level reflects the simple safety benefit that AVs bring to their occupants by diminishing the influence of certain contributing factors (e.g., distracted driving). At the second level, AVs could take evasive maneuvers to avoid a crash beyond what human drivers could perform. Both these levels of safety operate and are measured on the vehicle system level.

A third level of safety can be hypothesized and measured at the ecosystem level. AVs might reduce crashes by bringing more order and predictability to the broad roadway ecosystem. This would be particularly likely if and when there is a high level of communication among AVs, although this remains an uncertain prospect. When AVs reach a certain level of penetration into the vehicle fleet, they might influence roadway behavior beyond just the vehicles they are in immediate proximity to or contact with, elevating the safety of the entire fleet.[127]

An ecological study could be implemented in many ways. The simplest option is to compare a safety metric for location X to location Y, where X has AVs and Y does not. Alternatively, the safety metric for location X could be compared before or after AVs are available, in a time-series

[124] Morgenstern and Thomas, 1993.

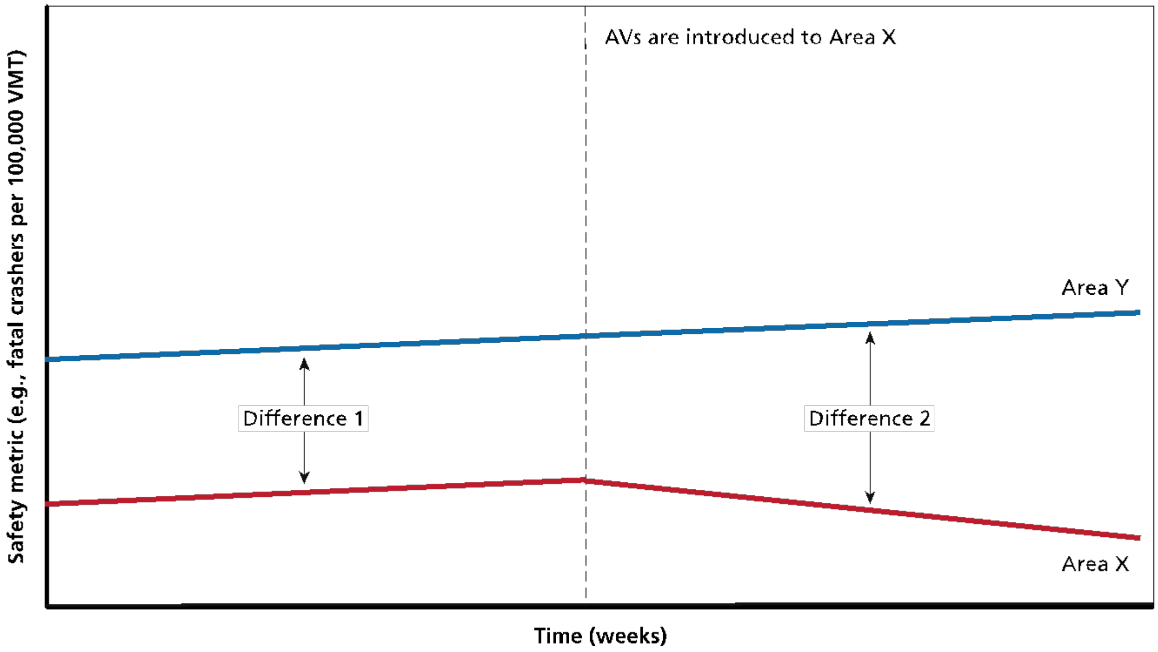
[125] Ecola et al., 2018.

[126] Ecological study design does not account for the degree to which individual conventional vehicles and AVs interact, nor does it account for the density of AVs in a geographic perimeter (Morgenstern and Thomas, 1993). This yields two of the biggest challenges in ecological studies: ecological bias and ecological fallacy. *Ecological bias* is likely to occur—there will be variation in the individual-level safety benefits because there is unlikely to be perfect mixing of AVs with conventional vehicles in geographic area X (Morgenstern and Thomas, 1993). An *ecological fallacy* could occur when conclusions are made about individuals based on group data; e.g., if area X has very good safety metrics, a given vehicle in that area is assumed to be very safe, when actually that one vehicle could be extremely unsafe and be bringing down the whole area's rate (Schwartz, 1994). Measures on the ecological level cannot be applied to the system level.

[127] Jane Lappin, email with authors, June 11–August 24, 2018.

design.[128] These approaches can be combined in the quasi-experimental difference-in-difference design (Figure 3.7).[129] Within this design, safety metrics from two or more areas are measured over time. After a period, AVs are introduced into one area. To measure the safety impact of AVs, Difference 2 (the difference between areas X and Y after AVs were introduced to area X) is compared with Difference 1 (the difference between areas X and Y before AVs were introduced). Additional complexities can be added, with multiple areas and cascading introduction of AVs, introduction of different types of AVs into different areas, considering variation over time, etc. The strength of the design comes from its built-in tolerance for the inherent differences between areas.[130] Such an analysis could occur prospectively or retrospectively.

Figure 3.7. A Difference-in-Difference Study Design



[128] Morgenstern and Thomas, 1993.

[129] Wing, Simon, and Bello-Gomez, 2018.

[130] Wing, Simon, and Bello-Gomez, 2018.

4. Conclusion

The approaches to measuring AV safety outlined in Chapter 3 are intended to be flexible but rigorous, offering commonality for AV developers and other stakeholders, from policymakers to safety advocates. This chapter enumerates additional considerations and offers some recommendations aimed at furthering broad public dialogue about the safety of vehicles that are highly (and eventually fully) automated.

AVs in the Broad Motor Vehicle Context

Beyond passenger vehicles, automation is of interest for trucks and buses. Although many of the technology development and safety measurement concepts covered here would apply, both the actual safety and the perception of safety of those vehicles are affected by their sheer size and mass. They also operate under regulations and with oversight that are different from those for passenger vehicles. Comparatively small automated cousins to trucks are also being developed for freight transport or delivery without human passengers (which means that safety concerns focus on other roadway users),[131] and comparatively small automated shuttle-buses already have been developed to transport small numbers of people (e.g., Navya).[132]

Meanwhile, the roadway environment itself is evolving in ways that can support both effective automation and safety. Gradually, more sensors are being embedded in road surfaces and connected to traffic management systems, and there are continuing efforts to develop systems that support communications between vehicles and roadway infrastructure. Communication among vehicles is expected to grow; it is already a feature in early efforts to support truck platooning, assisting with traffic flow and efficiency. V2X adds to both the mechanisms that can support AV safety and the complexity of the overall ecosystem.

AV Safety and Business Model

Generally, two distinct business models have dominated discussion of AVs: (1) consumer ownership and (2) fleets (including different taxi or ride-share services and institutional-support models) owned and operated by a single entity. Consumer ownership involves individuals who could vary in their abilities to maintain a vehicle's sensing, seeing, and operating systems appropriately.[133] For example, whereas a clean exterior might be a matter of aesthetics for a conventional vehicle, it could be essential for the proper performance (and thus, safety) of an AV

[131] Haddon, 2018; Higgins, 2018.

[132] Hawkins, 2018.

[133] Ecola et al., 2018.

that depends on cameras and other external sensors. Consumer owners, and the businesses they use that focus on automotive service and repair, might require education and support in sensor and camera maintenance. Fleet owners face similar challenges, but, with a large number of vehicles to support, they could choose to employ dedicated maintenance professionals.

Potential for Common Research and Development Infrastructure

The industry appears to be ambivalent about whether, when, where, and how to coordinate on key aspects related to development (see the next section). Efforts exist to promote testing in a common environment (e.g., the closed course of the American Center for Mobility[134]); so do efforts to promote development and use of common scenarios (e.g., the PEGASUS project), common testing code and procedures (e.g., Voyage's Open Autonomous Safety Initiative[135]), and common high-definition mapping and associated simulation (e.g., Baidu's Apollo platform[136]).

In brief, there seems to be more talking about sharing than actual sharing. But early efforts exist; for example, Vires recently transferred its format for describing road networks (OpenDRIVE) to the Association for Standardization of Automation and Measuring Systems (a German car manufacturer association) to promote broader use.[137] As has been noted, the U.S. Department of Transportation designated ten closed courses (out of 60 candidates) in 2017 to encourage both AV development and information-sharing.[138] Sharing of facilities and infrastructure might appeal most to those who cannot afford to develop and operate their own facilities and infrastructure.

Data-Sharing and Implications of Data as a Competitive Asset

AVs produce huge amounts of data, and questions have been raised about who beside the companies that produce or own AVs should have access to those data. There is little data-sharing among developers, nor is there much between developers and regulators or researchers. This is largely the result of the highly proprietary nature of AV development, variation in technologies that collect and process data, and the immense value of data as an asset and competitor differentiator,[139] combined with the lack of retail availability. Contents of California's mandatory disengagement reports show a lack of uniform reporting by developers, as discussed in Chapter 3. When crashes occur, state and federal entities undertake investigations that hinge on

[134] American Center for Mobility, undated.

[135] Voyage, undated-a.

[136] Apollo, undated.

[137] Association for Standardization of Automation and Measuring Systems, undated.

[138] U.S. Department of Transportation, 2017.

[139] One illustration of the costliness of not sharing comes from Scale API, which assists with labeling data for multiple AV players who might well pay for repeated but separate labeling of the same data. Marshall, 2018.

the willingness of developers to share and interpret data and other information not otherwise available.

This situation has led some observers to call for more sharing of data,[140] and to government explorations of the situation.[141] What data should be shared, with whom, at what level of detail, and how?[142] Types of data collected by AVs vary widely, although broad categorization features the following five items:

1. the external environment encountered by the vehicle
2. how the system perceived the external environment and processed resultant information
3. how the system interacted with the external environment
4. safety lagging or leading events
5. cybersecurity incidents (which will be discussed in next section).[143]

Theoretically, all data could be shared between companies and with government. However, it is extremely unlikely that items 2 and 3 would be shared. These data are particular to a system, and masking proprietary information while maintaining sufficient detail depth for the data to be useful verges on impossible. Companies view such data as proprietary assets. Additionally, sharing sensitive data requires strong security, which public repositories might not provide.

Regarding item 1, a data repository could contain the environmental circumstances encountered by the AV or specific environments, infrastructure, and traffic patterns that challenged the AV system.[144] Data would have to be anonymized. Such data could feed into scenarios used during development and demonstration and into scenario catalogs discussed in Chapter 3.[145] Exposing AVs in simulation or closed courses to such scenarios based on real events that are otherwise unlikely to be encountered by the AV system during development and demonstration makes the system more robust.[146]

Regarding item 4, existing crash databases and crash investigations currently provide data, albeit limited in detail in the former and limited in generalizability in the latter. Traffic safety statistics from these databases are publicly available,[147] and commentaries about specific vehicle

[140] Bryant Walker Smith has predicted that "when an automated vehicle developer shares its safety philosophy with the public through data and analysis . . . automated driving will be truly imminent." Smith, 2016.

[141] The U.S. Department of Transportation has been exploring data-sharing issues by convening stakeholders and discussing a possible framework categorizing different types of data and sharing options (U.S. Department of Transportation, undated) and of draft principles (U.S. Department of Transportation, 2018b). Feedback has included both the preference for data exchange to be voluntary and a recognition of the need for data standards (U.S. Department of Transportation, 2018a).

[142] U.S. Department of Transportation, 2018b. Principles for voluntary data-sharing were articulated by the U.S. Department of Transportation in mid-2018.

[143] U.S. Department of Transportation, 2017, 2018b.

[144] U.S. Department of Transportation, 2017.

[145] U.S. Department of Transportation, 2017.

[146] Koopman, 2018.

[147] Insurance Institute for Highway Safety, Highway Loss Data Institute, 2017b; NHTSA, 2018d; National Safety Council, undated-b.

models are available from such organizations as Consumers Union and the Insurance Institute for Highway Safety,[148] which have a history of buying, evaluating, and rating automobiles. A new data set containing detailed crash reconstructions for all lagging-measure events at or above a given severity would enhance understanding of how AVs fail and what happens when they do. Such data must be anonymized to protect road users' privacy. Within statistical databases, anonymization will become easier as more events occur. As is the case with existing databases, such data would not come from developers but from police reports, medical records, crash reconstructionist reports, etc.

An additional consideration regarding item 4, beyond the lack of detailed data, is that identifying AVs would pose a challenge when using traditional crash data sets (such as FARS and NASS/CDS) to understand AV safety. One option would be for vehicle identification numbers to include an indication of the highest level of automation at which any system on the vehicle is designed to operate.

Beyond vehicle safety, AVs collect broader information on the roadway ecosystem that could be useful to researchers and policymakers. Such data can identify problematic intersections, show the results of an infrastructure intervention, etc. Although those data are collected privately, it might be possible to achieve mutually beneficial sharing agreements among government, quasigovernment, and research organizations and companies. For example, AVs could provide information on road condition on a nonexclusive basis to a city department of transportation in exchange for real-time roadworks updates.[149]

Cybersecurity as a Factor

Cybersecurity bears importantly on safety. Today, even a nonautomated car is a network of computers on wheels. AVs are particularly vulnerable to cybersecurity attacks because they are made of computer-based systems—in technical jargon, AVs are cyber-physical systems or systems that embed computer-based elements in something that operates or interacts in the physical world. The transition toward AVs will only increase reliance on inevitably vulnerable hardware and software, both susceptible to cybersecurity problems. Recent experience by banks, utilities, and even the federal government suggest that successful cyberattacks are likely. The automotive industry's relative inexperience in addressing this risk also suggests caution should be exercised.

Beginning with the early use of computer-based systems and motor vehicles, automotive cybersecurity has been addressed separately from safety, attracting its own specialists and teams.

[148] See, for example, Consumer Reports, 2018.

[149] Laris, 2018.

It is the focus of standard-setting efforts and even conferences separate from those focused on safety.[150]

Cybersecurity for AVs will be further complicated as communications from the outside grow. Many concepts of AV operation rely at least in part on some kind of connectivity with other vehicles, infrastructure, or the internet. This connectivity is probably necessary, but it also increases cybersecurity risks. Even for conventional vehicles, over-the-air communication could (and to some extent already does) take place for diagnostic or updating purposes.[151]

The risks of a cybersecurity failure are considerable. At the lowest level, unsophisticated actors could cause simple vandalism by preventing the operation of parts of the vehicle, with safety consequences at varying levels of severity (e.g., preventing the opening of windows, interfering with the operation of brakes). Ransomware attacks, in which a perpetrator demands compensation (increasingly using anonymous cryptocurrencies) in return for restoring control of a software system, could be a viable business model for some criminal groups, which might be in other countries outside the easy reach of U.S. law enforcement. Relatively detailed personal data (including audio, video, and location data) could be collected and exploited. Large-scale terrorist or foreign nation-state attacks exploiting the same software or hardware vulnerabilities and utilizing numerous vehicles to attack critical infrastructure could cause mass casualties or sow panic.

While cybersecurity is of critical concern, it is distinct enough from other safety risks that it deserves separate treatment elsewhere.[152] For many safety scenarios, risk relates to inattention, mechanical or electronic threats, or negligence. In the case of cybersecurity, it is an actual adversary deliberately trying to exploit particular vulnerabilities. This is much harder to prevent and harder to detect. A hacker will probably not reveal that she has found a critical vulnerability until after that vulnerability is exploited and the damage is done.

While a thorough treatment of cybersecurity is outside the scope of this project, there are some obvious steps that can be taken by manufacturers and policymakers. Providing multiple and layered approaches to protecting a system (redundancy in its construction and defense in depth in its construction and operation) is critical for both safety and cybersecurity. A cyber-physical attack requires many steps to succeed, and efforts to prevent the attack at every stage are probably wise. Automakers should continue to share best cyber practices.[153] Given the national security risks involved, the probable inability of the civil justice system to create

[150] An Automotive Cybersecurity Standard (ISO/SAE AWI 21434) is under development focuses on cybersecurity engineering for road vehicles (Barber, 2018). A notable conference example is escar (undated), which focuses on embedded cybersecurity in cars. Cybersecurity conferences (e.g., USENIX-Security) also include papers on vulnerabilities, threats, and attacks on automotive systems.

[151] Consumer Reports has weighed in on this issue. Barry, 2018.

[152] For a discussion of how AV cybersecurity lapses might implicate liability, see Winkelman et al., forthcoming.

[153] Encouraged by the Department of Homeland Security, the Automotive Information Sharing and Analysis Center (undated) is a vehicle for such sharing.

adequate incentives for the scale of risks,[154] and the historic role of other nation-states in cyber-attacks, there is a strong argument for the involvement of the federal government in both leading and requiring strong cybersecurity protections.

Similar to cybersecurity, AV safety can be affected through actions that would present no danger to a conventional vehicle. Individuals will want to try to interact with or prank an AV, perhaps for notoriety. Such people might alter or remove infrastructure-based markings that AVs use to navigate. A recent study showed that AVs could be fooled or confused by defaced road signs that would not confuse human drivers.[155] As AVs proceed in the roadway, they must be robust to the chaos that they will encounter—not just from the environment and the mistakes of other road users, but also from deliberate attempts, malevolent or otherwise, to interact in undesirable ways.

Residual Uncertainty

Uncertainty exists in terms of individual AV actions and reactions and in terms of AVs' broader influence on the roadway ecosystem. An AV's view of the world grows steadily but is intrinsically incomplete.[156] Even within an AV's ODD, there are uncertainties that should be acknowledged by policymakers and communicated to the public. And even if AVs were completely certain within their ODDs, the roadways are not static. New environments and road users will populate the roadway ecosystem, presenting ongoing challenges to AVs.

Experts refer to the problems posed by "edge" and "corner" cases—situations that deviate from what is expected to be the norm but need to be addressed by the engineers in the interest of safe operation. The development process involves working through anticipated and unanticipated circumstances that are encountered; there is residual uncertainty from the unanticipated and not yet encountered—the unknown unknowns.

Regulation and data-sharing can help in managing that residual uncertainty. Conformance to ISO standards is intended, in part, to reduce this very issue.

Comparisons with Aviation

Aviation safety is sometimes suggested as an arena for comparison as AVs evolve. It is clearly a domain in which there has been progress from comparatively low levels to high levels of safety. This domain also has experienced growing levels of automation, including challenges associated with human-machine interaction in the context of automation. Along with medical

[154] Winkelman et al., forthcoming.

[155] Evtimov et al., 2017.

[156] The process of learning by doing, combining different modes of development and testing to capture unanticipated circumstances and challenges, was documented for Waymo's AVs by *The Atlantic* (Madrigal, 2017).

devices, aviation has driven advances in software safety, some of which could benefit the development of AVs. In aviation, data have been collected on a confidential basis for analyses that support the industry as a whole.[157]

Unlike the field of AV development, the field of aviation features a small and comparatively stable set of aircraft producers; aviation is subject to more regulation, and government involvement in certifying aircraft and key components is more accepted by industry—the context and the culture are different from what is observed in AV development today. Additionally, the carriers in the aviation field have agreed to cooperate rather than compete on safety based on the belief that one aviation death hurts the whole industry, regardless of on whose plane the death occurred. That sense of shared fate has yet to become evident among AV developers.

Communicating with the Public

Public conversation about AV safety is complicated. Experience with conventional vehicles, safety for other modes of transportation, safety of other kinds of computer-based systems, and public health risks all underscore the challenges associated with cognitive biases, limited numeracy, and general uncertainty among the public. For example, it might not be widely understood that although an AV might be capable of braking if a person runs out in front of it, the basic physics of its weight and speed might preclude it from being able to stop in time to avoid hitting the person. This could be exacerbated by unrealistic claims of near perfection on the part of AV boosters. Even compliance with seatbelt requirements remains uneven, and perceptions that ADAS are ineffective or irritating diminishes consumer trust in those features.[158] The utilitarian argument that AVs are expected to lower the annual numbers of deaths and injuries from car crashes substantially can be offset or even outweighed in popular perception by a single AV crash, especially given the breathless news coverage of any AV incident.[159] Furthermore, the variation in safety approaches taken by different AV developers could also confuse consumers.[160] So, too, could the capabilities of features associated with SAE Levels 1–3, which presume human engagement while also inducing the kinds of complacency that can compromise safety.[161] U.S. Secretary of Transportation Elaine Chao has called upon the AV industry to “step up and educate the public” about this new technology.[162] But until that happens,

[157] Aviation Safety Reporting System’s website (undated) underscores that associated reporting is “Confidential. Voluntary. Non-punitive.” Other reporting channels for safety incidents in aviation include the Federal Aviation Administration Hotline (2018) and the National Transportation Safety Board (undated).

[158] Insurance Institute for Highway Safety, Highway Loss Data Institute, 2017a.

[159] This could be a form of *dread risk*, a particularly aversive concern.

[160] Jill Ingrassia, Managing Director, Government Relations & Traffic Safety Advocacy, AAA, remarks at the Automated Vehicles Symposium 2018, San Francisco, Calif., July 11, 2018.

[161] NHTSA has intervened to stop the marketing of an aftermarket product intended to quiet the warnings associated with Tesla’s autopilot system. NHTSA, 2018c.

[162] Chao, 2018.

a gulf is likely to persist between those who understand it intimately and everyone else. The lack of a single measure, as discussed in Chapter 3, makes AV safety discussions complicated.

Communication is also necessary to inform consumers about an AV's abilities and about expectations for the user. Among other topics, people using AVs need to understand the limitations of the ODD and what will occur if the AV exceeds its ODD, including what "fallback," "limp home," or failure mitigation will look like. (These could involve turning on hazard lights, pulling over to the side of a road, or stopping in place.) Any expectations of users or owners, such as in-vehicle monitoring, also have to be conveyed; the expectation with HAVs is that a user would not have to intervene in vehicle operation. Safety-centric, appropriately couched communications can help the consumer not over-rely or under-rely on AV technology. Experience with lower levels of automation does make clear that better communications are needed.

Risk with AVs is a particularly important topic for consumers to understand. Accurate communication about the risks of AVs to the public will help individuals make decisions about transportation. If, as widely expected, AVs are substantially safer than conventional vehicles, accurate and effective risk communication will help spur adoption. But if consumers expect perfection and AVs are imperfect despite being much safer than conventional vehicles, adoption might be dangerously slow—i.e., more lives will be lost as a result of the slowness of adoption.[163]

Effective risk communications could also reduce litigation in three ways. Most concretely, effective risk communications could defeat a failure-to-warn claim.[164] If a defendant automaker can identify a particular communication of a particular risk to a plaintiff, it is unlikely that such a claim would succeed. Second, by creating appropriate expectations, risk communications could reduce the likelihood of anger and betrayal that can spur injured victims to file a lawsuit. Third, and most importantly, effective risk communications can lead to the safer use of the vehicles. If there are particular circumstances that lead to even small amounts of increased risk, warning users of these risks could lead to increased vigilance and reduced crashes and therefore fewer occasions to sue.

Unfortunately, effective risk communication is difficult. Humans process quantitative information poorly and are much more likely to be influenced by anecdotes or stories.[165] Humans also evaluate risk unevenly, exhibiting less concern over familiar risks than ones that are

[163] Kalra and Groves, 2017.

[164] This refers to a product liability claim wherein the seller failed to provide adequate warning or instructions for safe usage.

[165] Fischoff, Brewer, and Downs, 2011, p. 53. A study found that medical journals did a poor job of providing risk statistic–centric communication that was accurate and that newspaper articles were even worse. Moynihan et al., 2000.

uncertain, uncontrollable, inequitable, or particularly dreaded.[166] At least at first, consumers are likely to consider AVs uncertain, and they might have a particular fear of a robotic car killing them—even if that is less likely than a conventional car killing them. But over time, as the risks from AVs or robotic vehicles becomes more familiar, this particular dread might wane.

Fortunately, there is a science of risk communications developed in medicine and public health.[167] A full explanation of how those risk communications principles could be applied to AVs is outside the scope of this project, but several best practices can be usefully summarized.

To adequately inform, communications must contain the information needed for effective decisionmaking, and users need to be able access that information and understand what they access.[168] Best practices include providing numeric likelihoods of absolute risks that keep time frames and denominators constant. Pictographs can help convey this information effectively.[169] Safety communications might not follow these practices currently. A survey indicated that consumers most prefer to learn about their vehicle and its technology from vehicle manuals, from the dealership either at delivery or during sales, or online. However, these sources do not entirely overlap with how consumers report that they currently learn, chiefly from vehicle manuals and through trial and error.[170]

Ideally, the risk communications process itself should be tested. First, the process should be developed. Next, it should be evaluated to ensure consistency of message and accurate implementation. Finally, outcome evaluation should show whether the communication reached its goals and that the risks were in fact, understood by the target audience.

Developers and regulators can also work to demystify AVs, as Chao noted. When elevators were initially developed, there was considerable public fear. Being supported by unseen machinery while riding in a box that had the potential of dropping one to one's doom likely elicited considerable fear. To address this fear, manufacturers staged demonstrations of their automatic safety brakes. In a similar spirit, a recent exploration centered in Boston and connecting to other cities around the world demonstrated the value of an "AV petting zoo," an explicit effort to introduce consumers to AVs.[171]

In the case of elevators (and perhaps even more important than the demonstrations), legislators passed laws to require regular elevator inspections. To this day, elevators in most states are inspected regularly.[172] From a pure cost-benefit risk prevention perspective, this is

[166] Fischoff, Brewer, and Downs, 2011, p. 46.

[167] Fischoff, Brewer, and Downs, 2011.

[168] Fischoff, Brewer, and Downs, 2011, p. 19.

[169] Fischoff, Brewer, and Downs, 2011, pp. 59–61.

[170] Abraham et al., 2017.

[171] One idea suggested by a working group at the Autonomous Vehicle Symposium 2018 is that the lead developer of an AV take the first post-deployment ride, in the back seat. City of Boston, 2017; World Economic Forum and the Boston Consulting Group, 2018.

[172] Safety inspections for conventional automobiles are not handled consistently across states. Some states do not currently have safety inspections.

probably irrational—elevator failures are, thankfully, exceedingly rare. But the elevator inspections might have helped reassure a queasy public. It is possible that regular AV inspections or a similar regulatory regime could similarly allay public fears, as could some kind of transparent reporting during demonstration

Recommendations

The AV community's approach to safety will play a critical role in determining the success and viability of this technology. This report, in addition to offering a framework for measuring AV safety, identifies key recommendations for promoting AV safety to support the field's long-term viability.

First, during development, regulators and the public should be concerned with whether the AV development process is safe for the public, not whether the AV is progressing or learning in a particular way. Infractions, measures of roadmanship (on public roads), and outcomes are meaningful during this stage (because they indicate a risk to the public during the development process).[173] Concomitantly, a formal definition of roadmanship is needed for the development, demonstration, and deployment stages.[174]

Second, demonstration represents a stage apart from development and deployment that can be used for benchmarking and communicating about safety (recognizing that there are limits to what can be shown absent hundreds of millions or more miles driven). Demonstration is undertaken by testing through simulation and on closed courses and public roads with safety drivers—the settings discussed in this report. In closed courses and on public roads, demonstration protocols must control for variability in safety drivers. A formal protocol for the demonstration process (which could apply to simulators, simulations, and scenarios) would facilitate comparisons across companies and evidence of safety to the public and policymakers. This might suggest a role for a third party or department of motor vehicles.[175]

Third, during development, demonstration, and early deployment, when sufficient exposure has not been accumulated to allow for statistically valid comparison of rates, outcomes (e.g., crashes or an absence of crashes) should be evaluated as case studies. Such treatment of events shows a balance between learning from the event to fullest extent possible and not making statistics-based safety determinations that are beyond what the data support.[176]

[173] This recommendation is discussed at greater length in the Chapter 3 sections "Frame 1. Setting" and "From Frames to Framework."

[174] This recommendation draws from Chapter 3's "Measure Category 2: Leading Measures."

[175] This recommendation draws from the Chapter 3 sections "Demonstration" and "From Frames to Framework."

[176] This recommendation draws from the Chapter 3 sections "Frame 3. Measures," "From Frames to Framework," and "Measures to Metrics."

Fourth, targeted data-sharing (both between companies and with government) provides an opportunity to improve safety across the industry and to negotiate to receive data (e.g., around roadworks) in return. Regarding outcome measures, a protocol for reporting to government entities could be codified in terms of measures, context, format, frequency, data security, governance, and other factors.[177]

Fifth, a formal taxonomy around ODD is needed. Such a taxonomy should specify how ODDs convey where, when, and under what circumstances the AV can operate. Being within the developer-specified ODD does not guarantee that the AV will not encounter scenarios beyond its capabilities. Minimal-risk conditions should also be included.[178]

Sixth, given the challenges of measuring safety where the system changes constantly and at irregular intervals, this report outlines two approaches to measuring safety in association with AV system upgrades, moving averages and weighted moving averages. More research is needed to enhance understanding of and improvement of such methodologies.[179]

The Bigger Picture

As has been discussed in other RAND-published work,[180] the potential to minimize traffic fatalities by the middle of the 21st century hinges on progress from new technologies, policies, and other actions that promote a safety culture. The great race to develop safe and practical AVs should contribute to the overall goal of safe motor vehicle environments. This report's framework is intended to assist in that process. At the same time, it is important to acknowledge that all components of the ecosystem—including the roadway environment, its varied users, and other factors that go beyond a defined ODD—play a role in determining the safety as experienced with a given AV in a given set of circumstances.

Although the rise of AVs has been a story of disruption, there are already signs of maturation, at least on the industry front, drawing on connections between conventional and AV production and other industries. As a result, there is hope of more collective action among competitors—what some might call coopetition. A variety of consortia have emerged following a history of similar activity among conventional manufacturers,[181] there is broad participation in

[177] This recommendation draws from the section on "Data-Sharing and Implications of Data as a Competitive Asset" earlier in this chapter. Considerations around sharing scenarios can be found in Chapter 3's " A Discussion of the Setting Frame."

[178] This recommendation draws from Chapter 3's "Operational Design Domain." Comparisons involving AVs should include consideration of ODD, as discussed in the Chapter 3 sections "Operational Design Domain" and "Creating Comparisons."

[179] This recommendation draws from Chapter 3's "Upgrades and Updates."

[180] Ecola et al., 2018.

[181] For example, AV consortia include Self-Driving Coalition for Safer Streets (undated) and the Partnership for Transportation Innovation and Opportunity (undated). An example of a conventional vehicle consortia is Auto Alliance (undated).

relevant standard-setting activities, and a basis might be building for considering some greater degree of information-sharing about practices, tools, and even data. The European PEGASUS project was designed to foster sharing of scenarios that could be used for AV testing, although it reportedly encountered more reluctance to share than anticipated. More recently, the startup Voyage has announced that it would make safety-testing tools open source (although it remains to be seen how much others contribute),[182] and Baidu has developed a platform for sharing certain kinds of tools and data with an expectation that obtaining content requires also contributing content.[183] These organizations encapsulate an ongoing process of exploring where collaboration seems easy and where it seems difficult.

Although developers, regulators, and the public want AVs to be safer than conventional vehicles, the true impact of AVs is currently unknown. There have been early comparisons of AV driving data to non-AV driving data,[184] but interpreting these early results requires care and caution because of differences in severity threshold and a lack of generalizability beyond the geographies, time periods, and manufacturers included. Additionally, results associated with the AVs reflect the combination of the AV system and its supervising safety driver. Simulations of the entire roadway system could be done, but one can only speculate at this point. Developers and regulators must monitor the ongoing trends and shifts in the epidemiological profile of crashes and continue to refine measurement definitions and tools.[185] Only time, continued development, demonstration, and experience post-deployment will determine how safe AVs are and how safe they can become.

[182] Voyage, undated.

[183] Apollo, undated.

[184] Schoettle and Sivak, 2015; Teoh and Kidd, 2017; Blanco et al., 2016.

[185] An epidemiological profile describes the scope, severities, sociodemographics, risk factors, and other aspects of crashes. A sample epidemiological profile can be found at Centers for Disease Control and Prevention and Health Resources and Services Administration (2014).

Bibliography

Abraham, Hillary, Bryan Reimer, Bobbie Seppelt, Craig Fitzgerald, Bruce Mehler, and Joseph F. Coughlin, *Consumer Interest in Automation: Preliminary Observations Exploring a Year's Change*, Cambridge, Mass.: Massachusetts Institute of Technology AgeLab, white paper, May 25, 2017. As of September 5, 2018:
http://agelab.mit.edu/sites/default/files/
MIT%20-%20NEMPA%20White%20Paper%20FINAL.pdf

Almklov, Petter G., Ragnar Rosness, and Kristine Størkersen, "When Safety Science Meets the Practitioners: Does Safety Science Contribute to Marginalization of Practical Knowledge?" *Safety Science*, Vol. 67, 2014, pp. 25–36.

American Center for Mobility, homepage, undated. As of September 5, 2018:
http://www.acmwillowrun.org

Amersbach, Christian, and Hermann Winner, *Functional Decomposition: An Approach to Reduce the Approval Effort for Highly Automated Driving*, Darmstadt, Germany: Institute of Automotive Engineering, 2017.

Anderson, James M., Nidhi Kalra, Karlyn D. Stanley, Paul Sorensen, Constantine Samaras, and Tobi A. Oluwatola, *Autonomous Vehicle Technology: A Guide for Policymakers*, Santa Monica, Calif.: RAND Corporation, RR-443-2-RC, 2016. As of August 31, 2018:
https://www.rand.org/pubs/research_reports/RR443-2.html

Apollo homepage, undated. As of September 4, 2018:
http://apollo.auto

Arai, Yuji, Tetsuya Nishimoto, Yukihiro Ezaka, and Kenichi Yoshimoto, "Accidents and Near-Misses Analysis by Using Video Drive-Recorders in a Fleet Test," in *Proceedings of the 17th International Technical Conference on the Enhanced Safety of Vehicles (ESV) Conference*, Amsterdam, June 4–7, 2001. As of September 4, 2018:
https://pdfs.semanticscholar.org/609b/99dbb5787267d7c6f1147206e1bb49731bf3.pdf

Association for the Advancement of Automotive Medicine, "Overview," webpage, undated. As of September 4, 2018:
https://www.aaam.org/abbreviated-injury-scale-ais

Association for Standardization of Automation and Measuring Systems, "Kick-Off Workshop ASAM OpenDRIVE," webpage, undated. As of September 2, 2018:
https://www.asam.net/conferences-events/detail/kick-off-workshop-asam-opendrive

Auto Alliance homepage, undated. As of September 4, 2018: https://autoalliance.org

Automotive Information Sharing and Analysis Center, homepage, undated. As of September 5, 2018: https://www.automotiveisac.com

Aven, Terie, "What Is Safety Science?" *Safety Science*, Vol. 67, 2014, pp. 15–20.

Aviation Safety Reporting System, homepage, undated. As of September 5, 2018: https://asrs.arc.nasa.gov

Banerjee, Subho S., Saurabh Jha, James Cyriac, Zbigniew T. Kalbarczyk, and Ravishankar K. Iyer, *Hands Off the Wheel in Autonomous Vehicles? A Systems Perspective on Over a Million Miles of Field Data*, Urbana, Ill.: University of Illinois at Urbana-Champaign, 2018.

Barber, Angela, "Status of Work in Process on ISO/SAE 21434 Automotive Cybersecurity Standard," presentation, ISO SAE International, April 10, 2018. As of September 5, 2018: https://www.sans.org/summit-archives/file/summit-archive-1525889601.pdf

Barry, Keith, "Automakers Embrace Over-the-Air Updates, but Can We Trust Digital Car Repair?" *Consumer Reports*, April 20, 2018. As of September 5, 2018: https://www.consumerreports.org/automotive-technology/automakers-embrace-over-the-air-updates-can-we-trust-digital-car-repair

Blanco, Maria, Jon Atwood, Sheldon Russell, Tammy Trimble, Julie McClafferty, and Miguel Perez, *Automated Vehicle Crash Rate Comparison Using Naturalistic Data*, Blacksburg, Va.: Virginia Tech Transportation Institute, 2016.

Boehm, Barry, "Verifying and Validating Software Requirements and Design Specifications," IEEE Software, January 1984, pp. 75–88.

Bosch Engineering, "Validation of Highly Automated and Autonomous Automobile Systems," briefing slides, Frankfurt, Oberhursel: Automotive Safety and Security Week Testing ADAS and Self-Driving Cars, 2017.

Bryans, Jeremy, "From Automotive Safety to Automotive Security: Progress, Possibilities, and Pitfalls," briefing slides, IPC, undated.

BusinessDictionary, "safety," webpage, undated. As of September 2, 2018: http://www.businessdictionary.com/definition/safety.html

California Code of Regulations, Title 13, Motor Vehicles, Division 1, Department of Motor Vehicles, Chapter 1, Department of Motor Vehicles, Article 3.7, Testing of Autonomous Vehicles, Section 227.46, Reporting Disengagement of Autonomous Mode, undated. As of September 4, 2018: https://www.dmv.ca.gov/portal/wcm/connect/d48f347b-8815-458e-9df2-5ded9f208e9e/adopted_txt.pdf?MOD=AJPERES

California Department of Motor Vehicles, *Annual Report of Autonomous Vehicle Disengagements,* undated.

———, "Autonomous Vehicle Disengagement Reports 2017," webpage, 2017. As of September 4, 2018: https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/disengagement_report_2017

Centers for Disease Control and Prevention and Health Resources and Services Administration, *Integrated Guidance for Developing Epidemiologic Profiles: HIV Prevention and Ryan White HIV/AIDS Programs Planning*. Atlanta, Georgia, 2014. As of September 4, 2018: https://www.cdc.gov/hiv/pdf/guidelines_developing_epidemiologic_profiles.pdf

Chao, Elaine, U.S. Secretary of Transportation, remarks at the Automated Vehicles Symposium 2018, San Francisco, Calif., July 10, 2018.

Chen, Wenjun, Peter Cooper, and Mario Pinili, "Driver Accident Risk in Relation to the Penalty Point System in British Columbia," *Journal of Safety Research*, Vol. 26, No. 1, 1995, pp. 9–18.

City of Boston, "Robot Block Party," webpage, October 15, 2017. As of September 11, 2018: https://www.boston.gov/calendar/robot-block-party

Consumer Reports, "Cars with Advanced Safety Systems." June 29, 2018. As of September 5, 2018: https://www.consumerreports.org/car-safety/cars-with-advanced-safety-systems

Defense Advanced Research Projects Agency, "The DARPA Grand Challenge: Ten Years Later," webpage, March 13, 2014. As of September 2, 2018: https://www.darpa.mil/news-events/2014-03-13

Ecola, Liisa, Steven W. Popper, Richard Silberglitt, and Laura Fraade-Blanar, *The Road to Zero: A Vision for Achieving Zero Roadway Deaths by 2050*, Santa Monica, Calif.: RAND Corporation, RR-2333-NSC, 2018, As of July 24, 2018: https://www.rand.org/pubs/research_reports/RR2333.html

Els, Peter, "The First Level-3 Automated Vehicle Is on the Road: Is ISO Functional Safety and Analysis in Step?" *Automotive IQ*, April 25, 2018. As of April 26, 2018: https://www.automotive-iq.com/autonomous-drive/articles/first-level-3-automated-vehicle-road-iso-functional-safety-and-analysis

escar (Embedded Security in Cars), homepage, undated. As of September 5, 2018: https://www.escar.info

Every, Joshua L., Frank Barickman, John Martin, Sughosh Rao, Scott Schnelle, and Bowen Weng, *A Novel Method to Evaluate the Safety of Highly Automated Vehicles*, presented at the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration, Detroit, Michigan, 2017.

Evtimov, Ivan, Kevin Eykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Prakash, Amir Rahmati, and Dawn Song, "Robust Physical-World Attacks on Deep Learning Models," *Ground AI*, Vol. 1, 2017. As of September 6, 2018: https://www.groundai.com/project/robust-physical-world-attacks-on-deep-learning-models/

Federal Aviation Administration, *Safety Risk Management Policy*, Washington, D.C.: U.S. Department of Transportation, Order 8040.4B, May 2, 2017.

———, "FAA Hotline Reporting Form," August 13, 2018. As of September 5, 2018: https://hotline.faa.gov

Federal Highway Administration, *Surrogate Safety Assessment Model and Validation: Final Report*, McLean, Va.: U.S. Department of Transportation, February 2008. As of September 8, 2018:
https://rosap.ntl.bts.gov/view/dot/35896

Federal Ministry of Transport and Digital Infrastructure, *Ethics Commission Automated and Connected Driving Report*, Berlin, Germany, June 2017. As of September 2, 2018: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf

Felton, Ryan, "California's Autonomous Car Reports Are the Best in the Country—But Nowhere Near Good Enough," Jalopnik, February 1, 2018. As of September 4, 2018: https://jalopnik.com/californias-autonomous-car-reports-are-the-best-in-the-1822606953

Fischhoff, Baruch, Noel T. Brewer, and Julie S. Downs, eds., *Communicating Risks and Benefits: An Evidence-Based User's Guide*, Silver Spring, Md.: Department of Health and Human Services, Food and Drug Administration, August 2011.

Fischhoff, Baruch, Paul Slovic, Sarah Lichtenstein, Stephen Read, and Barbara Combs, "How Safe is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits," *Policy Sciences*, Vol. 9, No. 2, 1978, pp. 127–152.

Fraade-Blanar, Laura, and Nidhi Kalra, *Autonomous Vehicles and Federal Safety Standards: An Exemption to the Rule?* Santa Monica, Calif.: RAND Corporation, PE-258-RC, 2017. As of August 31, 2018: https://www.rand.org/pubs/perspectives/PE258.html

Frola, F. R., and C. O. Miller, *System Safety in Aircraft Management*, Washington D.C.: Logistics Management Institute, 1984.

Gordis, Leon, *Epidemiology*, 5th ed., New York: Elsevier, 2014.

Government of Alberta, *Leading Indicators for Workplace Health and Safety: A User Guide*, 2015. As of September 4, 2018: http://work.alberta.ca/documents/ohs-best-practices-BP019.pdf

Grim, Patrick, Robert Rosenberger, Adam Rosenfeld, Brian Anderson, and Robb E. Eason, "How Simulations Fail," *Synthese*, Vol. 190, No. 12, 2013, pp. 2367–2390.

Haddon, Heather, "Kroger Plans to Introduce Driverless Grocery Deliveries," *Wall Street Journal*, June 28, 2018. As of September 4, 2018: https://www.wsj.com/articles/kroger-plans-to-introduce-driverless-grocery-deliveries-1530190801

Haddon, William, Jr., "Advances in the Epidemiology of Injuries as a Basis for Public Policy," *Public Health Reports*, Vol. 95, No. 5, 1980, pp. 411–421.

———, *Approaches to Prevention of Injuries*, paper presented at the American Medical Association Conference on Prevention of Disabling Injuries, Miami, Fla., 1983.

Hartas, Dimitra, *Educational Research and Inquiry: Qualitative and Quantitative Approaches*, London: Bloomsbury Publishing, 2015.

Hawkins, Andrew J., "Senior Citizens Will Lead the Self-Driving Revolution," The Verge, January 10, 2018. As of September 4, 2018: https://www.theverge.com/2018/1/10/16874410/voyage-self-driving-cars-villages-florida-retirement-communities

Hayward, John C., "Near Miss Determination Through Use of a Scale of Danger," in *Proceedings of the 51st Annual Meeting of the Highway Research Board*, Washington, D.C., 1972, pp. 24–35.

Higgins, Tim, "Pizza Delivery Gears Up for Driverless Era," *Wall Street Journal*, June 26, 2018. As of September 4, 2018: https://www.wsj.com/articles/pizza-delivery-may-be-entering-a-new-era-1530029087

Hollnagel, Erik, "Is Safety a Subject for Science?" *Safety Science*, Vol. 67, 2014, pp. 21–24.

Hollnagel, Erik, Robert L. Wears, and Jeffrey Braithwaite, *From Safety-I to Safety-II: A White Paper*, Denmark: Resilient Health Care Net, published simultaneously by the University of Southern Denmark, University of Florida, and Macquarie University, 2015. As of September 2, 2018: https://www.england.nhs.uk/signuptosafety/wp-content/uploads/sites/16/2015/10/safety-1-safety-2-whte-papr.pdf

Hopkins, Andrew, "Issues in Safety Science," *Safety Science*, Vol. 67, 2014, pp. 6–14.

Ingrassia, Jill, Managing Director, Government Relations & Traffic Safety Advocacy, AAA, remarks at the Automated Vehicles Symposium 2018, San Francisco, Calif., July 11, 2018.

Insurance Institute for Highway Safety, Highway Loss Data Institute, "Lane Maintenance Systems Still a Turnoff for Many Drivers," *Status Report*, Vol. 52, No. 4, June 22, 2017a. As of September 5, 2018: https://www.iihs.org/iihs/sr/statusreport/article/52/4/3

———, "General Statistics: Yearly Snapshot—2016," webpage, December 2017b. As of September 11, 2018: https://www.iihs.org/iihs/topics/t/general-statistics/fatalityfacts/overview-of-fatality-facts

———, "Distracted Driving," webpage, May 2018. As of September 2, 2018: http://www.iihs.org/iihs/topics/t/distracted-driving/qanda

International Council on Systems Engineering, *Systems Engineering Handbook,* Version 3.1, August 2007.

———, *Systems Engineering Handbook: A Guide for System Lifecycle Processes and Activities*, 4th ed., Hoboken, N.J., Wiley Publishing & Sons, 2015.

International Organization for Standardization, homepage, undated-a. As of September 2, 2018: https://www.iso.org/standards.html

———, *Road Vehicles—Safety of the Intended Functionality,* Geneva, Switzerland, ISO 21448, undated-b. As of September 2, 2018: https://www.iso.org/standard/70939.html

———, *Road Vehicles—Functional Safety*, Part 1, *Vocabulary*, Geneva, Switzerland, ISO 26262, 2011. As of September 2, 2018: https://www.iso.org/obp/ui/#iso:std:iso:26262:-1:ed-1:v1:en

International Organization for Standardization, International Electrotechnical Commission, and Institute of Electrical and Electronics Engineers, *Systems and Software Engineering—Vocabulary*, ISO/IEC/IEEE 24765:2010(E), 2010.

ISO—*See* International Organization for Standardization.

Jha, Saurabh, Subho Sankar Banerjee, James Cyriac, Zbigniew Kalbarczyk, and Ravishankar K. Iyer, *AVFI, Fault Injection for Autonomous Vehicles*, Luxembourg, IEEE/IFIP International Conference on Dependable Systems and Networks, June 2018. As of September 2, 2018: https://www.researchgate.net/publication/325670464_AVFI_Fault_Injection_for_Autonomous_Vehicles

Johnsson, Carl, Aliaksei Laureshyn, and Tim De Ceunynck, “In Search of Surrogate Safety Indicators for Vulnerable Road Users: A Review of Surrogate Safety Indicators, *Transportation Reviews*, Vol. 38, No. 5, 2018. As of September 7, 2018: https://www.tandfonline.com/doi/pdf/10.1080/01441647.2018.1442888?needAccess=true&

Kalra, Nidhi, and David G. Groves, *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles*, Santa Monica, Calif.: RAND Corporation, RR-2150-RC, 2017. As of August 31, 2018: https://www.rand.org/pubs/research_reports/RR2150.html

Kalra, Nidhi, and Susan M. Paddock, Driving to Safety: *How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* Santa Monica, Calif.: RAND Corporation, RR-1478-RC, 2016. As of August 31, 2018: https://www.rand.org/pubs/research_reports/RR1478.html

Kim, Anita, David Perlman, Dan Bogard, and Ryan Harrington, *Review of Federal Motor Vehicle Safety Standards (FMVSS) for Automated Vehicles: Identifying Potential Barriers and Challenges for the Certification of Automated Vehicles Using Existing FMVSS*, Washington, D.C.: John A. Volpe National Transportation Systems Center, U.S. Department of Transportation, Preliminary Report, March 2016.

Klauer, Sheila G., Feng Guo, Bruce G. Simons-Morton, Marie Claude Ouimet, Susan E. Lee, and Thomas A. Dingus, “Distracted Driving and Risk of Road Crashes Among Novice and Experienced Drivers,” *New England Journal of Medicine*, Vol. 370, No. 1, 2014, pp. 54–59.

Koopman, Philip, *The Heavy Tail Safety Ceiling*, Greenville, S.C.: Automated and Connected Vehicle Systems Testing Symposium, June 20–21, 2018.

Koopman, Philip, and Michael Wagner, *Challenges in Autonomous Vehicle Testing and Validation*, SAE World Congress, 2016a.

———, “Autonomous Vehicle Safety: An Interdisciplinary Challenge,” *IEEE Intelligent Transportation Systems Magazine*, Vol. 9, No. 1, June 2016b, pp. 90–96.

———, *Toward a Framework for Highly Automated Vehicle Safety Validation*, Detroit, Mich.: 2018 SAE World Congress, SAE 2018-01–1071, April 10–12, 2018. As of September 2, 2018: https://users.ece.cmu.edu/~koopman/pubs/koopman18_av_safety_validation.pdf

Kriebel, David, Joel Tickner, Paul Epstein, John Lemons, Richard Levins, Edward L. Loechler, Margaret Quinn, Ruthann Rudel, Ted Schettler, and Michael Stoto, “The Precautionary Principle in Environmental Science,” *Environmental Health Perspectives*, Vol. 109, No. 9, 2001, pp. 871–876.

Lappin, Jane, email with authors, June 11–August 24, 2018.

Laris, Michael, “Federal Researchers Are Using Data from Waze and Maryland to Try to Predict Road Dangers,” *Washington Post*, 2018.

Le Coze, Jean-Chrisophe, Kenneth Pettersen, and Teemu Reiman, “The Foundations of Safety Science,” *Science Direct*, Vol. 67, 2014, pp. 1–5.

Leveson, Nancy, “A New Accident Model for Engineering Safer Systems,” *Safety Science*, Vol. 42, No. 4, April 4, 2004, pp. 237–270.

———, *Engineering a Safer World: Systems Thinking Applied to Safety*, Cambridge, Mass.: Massachusetts Institute of Technology Press, 2011. As of September 7, 2018: https://mitpress.mit.edu/books/engineering-safer-world

Lightstone, Amy S., Corinne Peek-Asa, and Jess F. Kraus, “Relationship Between Drivers Record and Automobile Versus Child Pedestrian Collisions,” *Injury Prevention*, Vol. 3, No. 4, 1997, pp. 262–266.

Loewenstein, George, “Out of Control: Visceral Influences on Behavior,” *Organizational Behavior and Human Decision Processes*, Vol. 65, No. 3, March 1996, pp. 272–292.

Lombaerts, Thomas, Stefan Schuet, Diana Acosta, and John Kaneshige, “On-Line Flight Envelope Determination for Impaired Aircraft,” in Thaddäus Baier and Matthias Heller, eds., *Robust Lateral Control of Future Small Aircraft*, New York: Springer, 2015, pp. 263–282. As of September 11, 2018: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20150022358.pdf

Lui, K.-J., and P. Marchbanks, “A Study of the Time Between Previous Traffic Infractions and Fatal Automobile Crashes, 1984–1986,” *Journal of Safety Research*, Vol. 21, No. 2, 1990, pp. 45–51.

Maddox, John, *Role of Proving Grounds in Three-Mode AV Validation Methodology*, paper presented at the Automated Vehicles Symposium, San Francisco, July 11, 2018.

Madrigal, Alexis C., “Inside Waymo’s Secret World for Training Self-Driving Cars,” *The Atlantic*, August 23, 2017. As of September 5, 2018: https://www.theatlantic.com/technology/archive/2017/08/inside-waymos-secret-testing-and-simulation-facilities/537648/

Mahmud, S.M. Sohel, Luis Ferreira, Shamsul Hoque, and Ahmad Tavassoli, "Application of Proximal Surrogate Indicators for Safety Evaluation: A Review of Recent Developments and Research Needs," *IATSS Research*, Vol. 41, No. 4, December 2017, pp. 153–163. As of September 7, 2018: https://www.sciencedirect.com/science/article/pii/S0386111217300286

Marshall, Aarian, "This Startup Would Like Self-Driving Car Companies to Share Data," *Wired*, August 7, 2018. As of September 5, 2018: https://www.wired.com/story/scale-ai-label-share-self-driving-data/

Merriam-Webster, "Safety," webpage, August 30, 2018. As of September 2, 2018: https://www.merriam-webster.com/dictionary/safety

Miller, Benjamin M., Frank Camm, Marjory S. Blumenthal, Jesse Lastunen, and Kenneth W. Miller, *Inching Toward Reform: Trump's Deregulation and Its Implementation*, Santa Monica, Calif.: RAND Corporation, PE-241-RC, 2017. As of September 2, 2018: https://www.rand.org/pubs/perspectives/PE241.html

Möller, Niklas, Sven Ove Hansson, and Martin Peterson, "Safety Is More Than the Antonym of Risk," *Journal of Applied Philosophy*, Vol. 23, No. 4, 2006, pp. 419–432.

Morgenstern, Hal, and Duncan Thomas, "Principles of Study Design in Environmental Epidemiology," *Environmental Health Perspectives*, Vol. 101, No. 4, 1993, pp. 23–38.

Moynihan, Ray, Lisa Bero, Dennis Ross-Degnan, David Henry, Kirby Lee, Judy Watkins, Connie Mah, and Stephen B. Soumerai, "Coverage by the News Media of the Benefits and Risks of Medications," *New England Journal of Medicine*, Vol. 342, No. 22, 2000, pp. 1645–1650.

Mynatt, Clifford R., Michael E. Doherty, and Ryan D. Tweney, "Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inference," *Quarterly Journal of Experimental Psychology*, Vol. 29, No. 1, 1977, pp. 85–95.

Nadimia, Navid, Hamid Behbahania, and Hamid Resa Shahbazib, 2016; "Calibration and Validation of a New Time-Based Surrogate Safety Measure Using Fuzzy Inference System," *Journal of Traffic and Transportation Engineering*, Vol. 3, No. 1, February 2016, pp. 51–58. As of September 7, 2018: https://www.sciencedirect.com/science/article/pii/S2095756415200153

National Highway Traffic Safety Administration, "Automated Vehicles for Safety," webpage, undated-a. As of August 31, 2018: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety#issue-road-self-driving

———, “Automatic Emergency Braking,” webpage, safercar.gov, undated-b. As of September 2, 2018: https://www.safercar.gov/Vehicle-Shoppers/Safety-Technology/AEB/aeb

———, “Event Data Recorder,” webpage, undated-c. As of September 2, 2018: https://www.nhtsa.gov/research-data/event-data-recorder

———, “OVSC Compliance Test Laboratories,” webpage, undated-d. As of September 2, 2018: https://one.nhtsa.gov/Vehicle-Safety/OVSC-Compliance-Test-Laboratories

———, “Regulations,” webpage, undated-e. As of September 2, 2018: https://www.nhtsa.gov/laws-regulations/fmvss

———, “Test Procedures,” webpage, undated-f. As of September 2, 2018: https://one.nhtsa.gov/Vehicle-Safety/Test-Procedures

———, “Test Specification Forms,” webpage, undated-g. As of September 2, 2018: https://one.nhtsa.gov/Vehicle-Safety/Test-Procedures/Test-Specification-Forms

———, *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*, Washington, D.C.: U.S. Department of Transportation, DOT HS 810 594, 2006. As of September 4, 2018: https://vtechworks.lib.vt.edu/bitstream/handle/10919/55090/DriverInattention.pdf

———, *Pre-Crash Scenario for Crash Avoidance Research*, Washington, D.C.: U.S. Department of Transportation, DOT HS 810 767, 2007. As of September 2, 2018: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/pre-crash_scenario_typology-final_pdf_version_5-2-07.pdf

———, *Safety-Related Defects and Automated Safety Technologies*, Washington, D.C.: U.S. Department of Transportation, NHTSA Enforcement Guidance Bulletin 2016-02, Docket No. NHTSA-2016-0040, 2016a.

———, “Accelerating the Next Revolution in Roadway Safety,” Washington, D.C., U.S. Department of Transportation, Federal Automated Vehicles Policy, September 2016b.

———, *Motor Vehicle Safety Defects and Recalls: What Every Vehicle Owner Should Know*, Washington, D.C.: U.S. Department of Transportation, DOT HS 808 795, August 2017a.

———, *Automated Driving Systems 2.0: A Vision for Safety*, Washington, D.C.: U.S. Department of Transportation, DOT HS 812 442, September 2017b. As of September 2, 2018: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

———, “Voluntary Safety Self-Assessment Template,” September 2017c. As of September 2, 2018: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/voluntary_safety_self-assessment_for_web_101117_v1.pdf

———, “Monthly Compliance Status Reports,” webpage, 2018a. As of September 2, 2018: https://icsw.nhtsa.gov/cars/testing/comply/monthly/

———, *Removing Regulatory Barriers for Vehicles with Automated Driving Systems*, Washington, D.C.: U.S. Department of Transportation, proposed rule, January 18, 2018b.

———, “Consumer Advisory: NHTSA Deems ‘Autopilot Buddy’ Product Unsafe,” press release, June 18, 2018c. As of September 5, 2018: https://www.nhtsa.gov/press-releases/consumer-advisory-nhtsa-deems-autopilot-buddy-product-unsafe

———, “Summary of Motor Vehicle Crashes,” fact sheet, August 2018d. As of September 11, 2018: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812580

National Institutes of Health, *Theory at a Glance: A Guide for Health Promotion Practice*, 2nd ed., Washington, D.C., 2005.

National Research Council, *Improving Risk Communication*, Washington, D.C.: National Academies, 1989.

———, Understanding Risk: Informing Decisions in a Democratic Society, Washington, D.C.: The National Academies Press, 1996.

National Safety Council, “Injury Facts,” webpage, undated-a. As of September 2, 2018: https://injuryfacts.nsc.org

———, “2017 Estimates Show Vehicle Fatalities Topped 40,000 for Second Straight Year,” webpage, undated-b. As of September 11, 2018: https://www.nsc.org/road-safety/safety-topics/fatality-estimates

National Transportation Safety Board, “Report an Aircraft Incident to the NTSB,” webpage, undated. As of September 5, 2018: https://www.ntsb.gov/Pages/Report.aspx

NHTSA—*See* National Highway Traffic Safety Administration.

Nowakowski, Christopher, Steven E. Shladover, and Ching-Yao Chan, “Determining the Readiness of Automated Driving Systems for Public Operation: Development of Behavioral Competency Requirements,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2559, 2016, pp. 65–72.

Nowakowski, Christopher, Steven E. Shladover, Ching-Yao Chan, and Han-Shue Tan, "Development of California Regulations to Govern Testing and Operation of Automated Driving Systems," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1, 2015, pp. 1–16.

Nuro, *Delivering Safety, Nuro's Approach*, Mountain View, Calif., 2018. As of September 20, 2018:
https://static1.squarespace.com/static/57bcb0e02994ca36c2ee746c/t/5b9a00848a922d8eaecf65a2/1536819358607/delivering_safety_nuros_approach.pdf

O'Neill, Brian, "Preventing Passenger Vehicle Occupant Injuries by Vehicle Design—A Historical Perspective from IIHS," *Traffic Injury Prevention*, Vol. 10, No. 2, 2009, pp. 113–126.

Occupational Safety and Health Administration, "The Business Case For Safety: Adding Value and Competitive Advantage," presentation, Georgetown University, March 2005. As of September 4, 2018:
https://www.osha.gov/dcsp/success_stories/compliance_assistance/abbott/abbott_casestudies/index.html

———, "The Use of Metrics in Process Safety Management (PSM) Facilities," fact sheet, October 2016. As of September 4, 2018:
https://www.osha.gov/Publications/OSHA3896.pdf

Organisation for Economic Co-operation and Development, *Safer Roads with Automated Vehicles*, Paris, France: International Transport Forum, 2018.

Partnership for Transportation Innovation and Opportunity, homepage, undated. As of September 4, 2018:
https://ouravfuture.org/

Pasztor, Andy, "NASA Safety Watchdogs Raise Concerns About SpaceX, Boeing Spacecraft," *Wall Street Journal*, January 12, 2018.

Peden, Margie, Richard Scurfield, David Sleet, Dinesh Mohan, Adnan A. Hyder, Eva Jarawan, and Colin D. Mathers, *World Report on Road Traffic Injury Prevention*, Geneva, Switzerland: World Health Organization, 2004.

PEGASUS Research Project, homepage, undated-a. As of September 2, 2018:
https://www.pegasusprojekt.de/en/home

———, "About," undated-b. As of September 2, 2018:
https://www.pegasusprojekt.de/en/about-PEGASUS

Perrow, Charles, *Normal Accidents: Living with High Risk Technologies*, revised edition, Princeton, N.J.: Princeton University Press, 1999.

Prescott, Al, Associate General Counsel for Tesla, "Autonomous Mode Disengagements for Reporting Year 2017: Tesla," memorandum to California Department of Motor Vehicles, December 31, 2017. As of September 4, 2018:
https://www.dmv.ca.gov/portal/wcm/connect/f965670d-6c03-46a9-9109-0c187adebbf2/Tesla.pdf?MOD=AJPERES&CVID=

Rahwan, Iyad, "Society-in-the-Loop: Programming the Algorithmic Social Contract," *Ethics and Information Technology*, Vol. 20, No. 1, 2018, pp. 5–14.

Renn, Ortwin, and Debra Levine, "Credibility and Trust in Risk Communication," in Roger E. Kasperson and Pieter Jan M. Stallen, eds., *Communicating Risks to the Public*, Boston: Kluwer, 1991, pp. 175–217.

Roose, Kevin, "Can Ford Turn Itself into a Tech Company?" *New York Times Magazine*, November 9, 2017. As of August 31, 2018:
https://www.nytimes.com/interactive/2017/11/09/magazine/tech-design-autonomous-future-cars-detroit-ford.html

Runyan, Carol W., "Using the Haddon Matrix: Introducing the Third Dimension," *Injury Prevention*, Vol. 21, No. 2, 1998, pp. 302–307.

SAE International, homepage, undated. As of August 31, 2018:
https://www.sae.org/

SAE Mobilus, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, Warrendale, Pa., J3016_201609, September 30, 2016. As of August 31, 2018:
https://saemobilus.sae.org/content/j3016_201609

———, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, Warrendale, Pa., J3016_201806, June 15, 2018. As of August 31, 2018:
https://saemobilus.sae.org/content/J3016_201806/

Savitz, Scott, Henry H. Willis, Aaron C. Davenport, Martina Melliand, William Sasser, Elizabeth Tencza, and Dulani Woods, *Enhancing U.S. Coast Guard Metrics*, Santa Monica, Calif.: RAND Corporation, RR-1173-USCG, 2015, As of May 25, 2018:
https://www.rand.org/pubs/research_reports/RR1173.html

Schoettle, Brandon, and Sivak, Michael, *A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles*, Ann Arbor, Mich.: University of Michigan Transportation Research Institute, October 2015.

Schwartz, Sharon, "The Fallacy of the Ecological Fallacy: The Potential Misuse of a Concept and the Consequences," *American Journal of Public Health*, Vol. 84, No. 5, 1994, pp. 819–824.

Self-Driving Coalition for Safer Streets, homepage, undated. As of September 4, 2018: http://www.selfdrivingcoalition.org

Shalev-Shwartz, Shai, Shaked Shammah, and Amnon Shashua, *On a Formal Model of Safe and Scalable Self-Driving Cars*, Jerusalem, Israel: Mobileye, 2017. As of September 4, 2018: https://arxiv.org/pdf/1708.06374.pdf

Slovic, Paul, *The Perception of Risk (Risk, Society, and Policy)*, London, England: Earthscan, 2000.

Smith, Aaron, and Monica Anderson, "Americans' Attitudes Toward Driverless Vehicles," in *Automation in Everyday Life*, Washington, D.C.: Pew Research Foundation, October 4, 2017. As of August 31, 2018: http://www.pewinternet.org/2017/10/04/americans-attitudes-toward-driverless-vehicles/

Smith, Bryant Walker, "The Public Safety Case," presentation, Law of the Newly Possible wiki, July 18, 2016. As of September 5, 2018: https://newlypossible.org/files/presentations/2016-07-18_PublicSafetyCase.pdf

"Standardization Efforts on Autonomous Driving Safety Barely Under Way," *Hansen Report on Automotive Electronics*, February 2017.

Statista, "Average Age of Passenger Cars in the U.S. from 1995 to 2016 (in years)," webpage, 2017. As of May 28, 2018: https://www.statista.com/statistics/261877/average-age-of-passenger-cars-in-the-united-states/

Stern, Paul C., and Harvey V. Fineberg, eds., *Understanding Risk: Informing Decisions in a Democratic Society*, Washington, D.C.: Committee on Risk Characterization, National Research Council, 1996.

Stilgoe, Jack, "We Need New Rules for Self-Driving Cars," *Issues in Science and Technology*, Vol. 34, No. 3, 2018, pp. 52–57.

Stipancic, Joshua, Luis Miranda-Moreno, and Nicolas Saunier, "Vehicle Manoeuvers as Surrogate Safety Measures: Extracting Data from the GPS-Enabled Smartphones of Regular Drivers," *Accident Analysis and Prevention*, Volume 115, June 2018, pp. 160–169. As of September 11, 2018: https://www.sciencedirect.com/science/article/pii/S000145751830109X

Teoh, Eric R., and David G. Kidd, "Rage Against the Machine? Google's Self-Driving Cars Versus Human Drivers," *Journal of Safety Research*, Vol. 63, 2017, pp. 57–60.

Tisca, Ionela Adriana, Nicolae Istrat, Constantin Dan Dumitrescu, and Georgica Cornu, "Issues Concerning the Road Safety Concept," *Procedia Economics and Finance*, Vol. 39, 2016, pp. 441–445. As of September 2, 2018:
https://www.sciencedirect.com/science/article/pii/S221256711630346X

Uchida, Nobuyuki, Maki Kawakoshi, Takashi Tagawa, and Tsutomu Mochida, "An Investigation of Factors Contributing to Major Crash Types in Japan Based on Naturalistic Driving Data," *IATSS Research*, Vol. 34, No. 1, 2010, pp. 22–30.

University of California PATH Program, *Peer Review of Behavioral Competencies for AVs*, Berkeley, California: University of California, Berkeley, 2016.

U.S. Code, Title 49, Transportation, Subtitle VI, Motor Vehicle and Driver Programs, Part A, General, Chapter 301, Motor Vehicle Safety.

U.S. Department of Defense, *System Safety*, Washington, D.C.: No. MIL-STD-8824, February 10, 2000.

U.S. Department of Transportation, "Draft AV Data Framework," webpage, undated. As of September 5, 2018:
https://www.transportation.gov/av/data/DraftAVDataFramework

———, "U.S. Department of Transportation Designates 10 Automated Vehicle Proving Grounds to Encourage Testing of New Technologies," webpage, January 19, 2017. As of September 2, 2018:
https://www.transportation.gov/briefing-room/dot1717

———, *Roundtable on Data For Automated Vehicle Safety*, Washington, D.C., summary report, 2018a.

———, "Data for Automated Vehicle Integration," webpage, September 4, 2018b. As of August 2, 2018:
https://www.transportation.gov/av/data

Voyage, "Collaborating for a Safe Autonomous Future," webpage, undated-a. As of September 5, 2018:
https://voyage.auto/open-autonomous-safety/

———, "Open-Sourcing Our Approach to Autonomous Safety," webpage, undated-b. As of September 4, 2018:
https://news.voyage.auto/open-sourcing-our-approach-to-autonomous-safety-434b1ab13a93

Wang, Chen, and Nikiforos Stamatiadis, "Evaluation of a Simulation-Based Surrogate Safety Metric," *Accident Analysis and Prevention*, Vol. 71, October 2014, pp. 82–92. As of September 11, 2018:
https://www.sciencedirect.com/science/article/pii/S0001457514001365

Waymo, *Report on Autonomous Mode Disengagements for Waymo Self-Driving Vehicles in California*, Mountain View, Calif.: Alphabet Inc., December 2016. As of September 4, 2018: https://www.dmv.ca.gov/portal/wcm/connect/946b3502-c959-4e3b-b119-91319c27788f/GoogleAutoWaymo_disengage_report_2016.pdf?MOD=AJPERES&CVID=%20%20%20Cf

———, *On the Road to Fully Self-Driving*, Mountain View, Calif.: Alphabet Inc., 2018. As of September 2, 2018: https://storage.googleapis.com/sdc-prod/v1/safety-report/Safety%20Report%202018.pdf

Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez, "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research," *Annual Review of Public Health*, Vol. 39, 2018, pp. 453–469.

Winkelman, Zev, Maya Buenaventura, James M. Anderson, Nahom Beyene, Pavan Katkar, and Greg Baumann, *When Hacked Autonomous Vehicles Do Damage, Who May Face Liability?* Santa Monica, Calif.: RAND Corporation, forthcoming.

Wolfram Language and System Documentation Center, "Heavy Tail Distributions," webpage, undated. As of September 2, 2018: http://reference.wolfram.com/language/guide/HeavyTailDistributions.html

World Bank, "Urban Transport Safety and Security," in *Cities on the Move*, Washington, D.C., August 2002. As of September 2, 2018: http://siteresources.worldbank.org/INTURBANTRANSPORT/Resources/chapter5.pdf

World Economic Forum and the Boston Consulting Group, *Reshaping Urban Mobility with Autonomous Vehicles: Lessons from the City of Boston*, Geneva: REF 140518, June 2018. As of September 11, 2018: http://www3.weforum.org/docs/WEF_Reshaping_Urban_Mobility_with_Autonomous_Vehicles_2018.pdf

Yoshida, Junko, "Uber Fatality Sends AVs Back to Safety 101," *EE Times*, July 19, 2018. As of September 6, 2018: https://www.eetimes.com/document.asp?doc_id=1333446

This report presents a framework for measuring safety in automated vehicles (AVs) that could be used broadly by companies, policymakers, and the public. In it, the authors considered how to define safety for AVs, how to measure safety for AVs, and how to communicate what is learned or understood about AVs. Given AVs' limited total on-road exposure compared with conventional, human-driven vehicles, the authors also consider options for proxy measurements—i.e., factors that might be correlated with safety— and explore how safety measurements could be made in simulation and on closed courses. The report identifies key concepts and illuminates the kinds of measurements that might be made and communicated. It presents a structured way of thinking about how to measure safety at different stages of an AV's evolution, and it proposes a new kind of measurement. While acknowledging that the closely held nature of AV data limits the amount of data that are made public or shared between companies and with the government, the report highlights the kinds of information that could be presented in consistent ways in support of public understanding of AV safety.

RAND JUSTICE, INFRASTRUCTURE, AND ENVIRONMENT

www.rand.org

$29.00

ISBN-10 1-9774-0164-3
ISBN-13 978-1-9774-0164-9
52900

9 781977 401649

RR-2662