PROGRAM ON U.S.-JAPAN RELATIONS
Weatherhead Center FOR INTERNATIONAL AFFAIRS
HARVARD UNIVERSITY

HARVARD UNIVERSITY

AUTOMATED VEHICLES AND THE TROLLEY PROBLEM

Keisuke Sakanoue

Harvard Program on U.S.-Japan Relations
Occasional Paper Series
22-SK

https://programs.wcfia.harvard.edu/us-japan/research

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AV	Automated vehicle
AD	Automated driving
ADS	Automated driving system
GM	General Motors
NHTSA	National Highway Traffic Safety Administration
ODD	Operational design domain
RAD	Reasonable alternative design
SAE	Society of Automotive Engineers International)
USDOT	U.S. Department of Transportation

Introduction

In recent years, with the rapid progress of science and technology, various technologies that were once thought to be only in realm of science fiction are now appearing in the real world. Automated vehicles (AVs) that do not require human operators are one of them. AVs are said to be the ultimate dream for automobiles, and with the advancement of science and technology and the growing needs of society, they are now expected to become a reality. As such, industry, academia, and government are working together to research and develop the technology, study the legal system, and improve the environment for AVs in countries around the world.

AVs enable safer and smoother driving than human driving by having an automated driving system (ADS) take over the cognitive, predictive, judgmental, and operational tasks necessary for safe driving of vehicles that have heretofore been performed by humans. This is expected to solve many social issues, such as reducing traffic accidents, easing traffic congestion, and securing means of transportation in depopulated areas. Concerning traffic accidents in particular, there were 2,636 traffic fatalities in Japan in 2021,[1] and, although the number has been decreasing in recent years, many precious lives are still lost. The corresponding number in the United States in 2020 was 38,680, the highest since 2007.[2] In addition, approximately 1.3 million lives are lost each year worldwide due to traffic accidents.[3] The United Nations General

[1] The National Police Agency of Japan website. "The number of traffic accident fatalities in 2021." 4 Jan. 2022, https://www.npa.go.jp/news/release/2022/20220104001jiko.html.
[2] The National Highway Traffic Safety Administration website. "Overview of motor vehicle crashes in 2020" Mar. 2022,
https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266.
[3] The World Health Organization. "Road traffic injuries" 21 Jun. 2021,
https://www.nhtsa.gov/press-releases/2020-fatality-data-show-increased-traffic-fatalities-during-pandemic.

Assembly has set a goal of halving the global number of deaths and injuries from road traffic accidents by 2030.[4]

The concept of automated driving (AD) itself is not new, and it is said that the world's first company to propose it was the American automaker General Motors (GM) some 80 years ago.[5,6] At the World Expo held in New York City in 1939 and 1940, GM introduced a diorama of a car running by an ADS at a booth called Futurama (a word made by combining "future" and "panorama"). In this booth, visitors sat in a "moving chair" that automatically rotated around the diorama. The diorama simulated a futuristic city of the 1960s, and the audience felt as if they were taking a scenic flight over it. In this diorama, a highway model for AD called the "Automated Highway" was proposed. In this highway, a vehicle guided by radio waves maintains the proper distances between vehicles and automatically drive to the destination according to the radio signals. A concept that would lead to an ADS with road-vehicle coordination, such as the one currently under consideration, was thus proposed. It is said that a total of about 25 million people visited this exhibition.

Since then, the momentum for the practical application of AD has been growing around the world as many countries have been conducting research on AD technologies and undertaking various demonstration tests. Most recently, in March 2021, Honda, a Japanese automobile company, launched the world's first owner AV equipped with an ADS that corresponds to Level 3 AD (see below for an explanation of AD levels.).[7,8] This system is called Traffic Jam Pilot, and

[4] Ibid.
[5] Furukawa, Osamu. "Development of automated driving technology: Its history and direction for practical application" Grand Prix Publishing Co. 2019.
[6] Hosaka, Akio, and Aoki keiji and Sadayuki tsugawa. "Automated driving (2nd Edition)-system structure and elemental technologies " Morikita Publishing Co. 2020.
[7] HONDA. "Launch of the new LEGEND equipped with Honda Sensing Elite" 4 Mar. 2021, https://www.honda.co.jp/news/2021/4210304-legend.html.
[8] Nikkei. "Honda to launch the world's first "\'Level 3' Automated driving system on May. 5" 4 Mar. 2021,

it operates the accelerator, brake, and steering wheel while monitoring the surrounding area on behalf of the driver under certain conditions during traffic jams on highways. The system is capable of driving, stopping, and restarting in the same lane, while maintaining a safe distance from the vehicle ahead, by changes in the vehicle's speed. This is said to reduce fatigue and stress during traffic jams by allowing drivers to operate the navigation system, such as watching a TV monitor or a DVD on the navigation screen and searching for destinations.

AVs are divided into various levels based on how much of the driving tasks the ADS takes on for the driver. Various classifications and definitions of AD have been proposed, but the Society of Automotive Engineers International (SAE) classification and definition (SAE J3016)[9] published in September 2016 is often used as for common understanding worldwide. This definition is also used in the research of the Autonomous Driving Subcommittee (ITS/AD), a subcommittee of the World Forum for Harmonization of Vehicle Regulations (WP29) of the United Nations Economic Commission for Europe.[10] According to SAE J3016, AD is classified into six levels, from 0 to 5. Of these, levels 3 to 5 are equipped with ADSs where the ADS (when activated) performs all of the dynamic driving tasks. For SAE Level 3, the ADS performs all dynamic driving tasks in the operational design domain (ODD), but the driver must respond to the system's intervention requests, if it is difficult to continue operation. For SAE Levels 4 and Level 5, the ADS will perform all dynamic driving tasks and responses to cases where continued operation is difficult, but for SAE Level 4, the ADS will operate in the ODD, while for SAE Level 5, the ADS will operate without domain limitations.

---

https://www.nikkei.com/article/DGXZQODZ032HJ0T00C21A3000000/.

[9] The SAE International J3016. "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles" Sep. 2016, https://www.sae.org/standards/content/j3016_201609/.

[10] Sekine, Michiaki, and Kaneo Hiramatsu. "Definition of automated driving technology and status of study on items for international standardization at the United Nations Autonomous Driving Subcommittee (ITS/AD)" Summary of the 2017 lecture at the Traffic Safety and Environment Research Institute Forum, 2017, pp. 47-50, https://www.ntsel.go.jp/Portals/0/resources/forum/2017files/1201_1430.pdf.

Currently, demonstration experiments and examination of legal issues are being conducted around the world for the practical application of AVs, and the trolley problem as a dilemma situation is increasingly being discussed with regard AD. In AVs of SAE Level 4 or higher, which are called highly AD or fully AD, humans do not operate the AVs within the ODD of the ADS, even when a situation arises that makes it difficult for the ADS to continue operating. While AVs are thought to be capable of significantly reducing the incidence of traffic accidents compared to vehicles driven by humans,[11] this does not mean that traffic accidents will be entirely eliminated.[12] If traffic accidents caused by AVs are expected to occur, it would be desirable to operate AVs in such a way as to cause less damage in the event of a traffic accident, and the following situations can be envisioned.

Suppose that an AV is driving, and there are five pedestrians in the path of the vehicle, and, even if the emergency brake is applied, a collision cannot be avoided. If the vehicle turns the steering wheel toward the sidewalk, the vehicle can avoid colliding with the pedestrians, but the vehicle may collide with one pedestrian on the sidewalk and cause a fatality.

In a case like this, the resulting damage would be minimal, but is such an algorithm for an ADS acceptable? Also, who is responsible for pedestrian fatalities in the event of an actual traffic accident? If a human being were to be driving a vehicle and find himself or herself in the situation described above, he or she would make decisions and take actions reflexively, and since humans do not have a complete grasp of the surrounding road traffic environment, the pros and cons of such decisions and actions would be questioned after the fact, depending on the

[11] The Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society website. "Outline of system development for automated driving" 17 Apr. 2019, http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20180413/auto_drive.pdf.

[12] For example, in a study on artificial intelligence, it was said that "If self-driving cars cut the roughly 40,000 annual U.S. traffic fatalities in half, the car makers might get not 20,000 thank-you notes, but 20,000 lawsuits." Russell, Stuart, and Daniel Dewey, Max Tegmark. "Research Priorities for Robust and Beneficial Artificial Intelligence." Ai Magazine, 36.4, 2015, pp.105-14.

individual and specific situation. In the case of AVs, however, they are driven according to a predetermined algorithm based on an understanding of the surrounding road traffic environment, so the pros and cons of the said predetermined algorithm would become an issue.

The trolley problem has been discussed for a long time as a thought experiment in ethics, "Is it permissible to actively sacrifice the one to save the many?" And it has been treated in recent years as a real problem with regard to the actualization of ADs. Regarding the trolley problem, the outcome of the choice depends on what kind of ethical thought process is taken. If we follow the utilitarian process, sacrificing one life to avoid the sacrifice of many is considered acceptable, while if we follow the duty theory, sacrificing one life to save many lives is not. It is very difficult to solve such a problem given the clash of different ethical views. Therefore, the trolley problem has been discussed as a thought experiment in ethics until now, but a situation has arisen in which it is assumed that some kind of answer will be needed as this is a real problem toward the realization of AD. Since the trolley problem is an unavoidable issue in Japan, where discussions on the realization of SAE Level 4 or higher AD are currently underway, investigating the status of the dialogues in the United States will help those in Japan.

Therefore, as a premise for discussion, this paper will first outline the governments' considerations and promotion of AD in Japan and the United States. After that, I would like to discuss the debates on the trolley problem in general and then the status of studies on the trolley problem related to AD in the United States.

Unless otherwise specified, the opinions expressed in this paper are the author's own and not the official views of the Japanese government.

Automated Vehicle Policies in Japan and the United States

Government Policies and Initiatives for Automated Vehicles in Japan

In Japan, the "Public-Private ITS Concept and Roadmap" was formulated in 2014 for planning and implementing strategies for the early realization of AD in cooperation with the public and private sectors, and it has been revised every year since then based on successive changes in circumstances.[13] The roadmap sets the government's goal of realizing an AD service with remote monitoring only (SAE Level 4) in limited areas by 2022 and nationwide deployment of the same service by 2025. The government also aims to realize SAE Level 4 automated trucks on expressways and SAE Level 4 AD services in spaces where human-driven vehicles and AVs are mixed, also by 2025.[14]

The realization of AD requires the systematic development of traffic-related laws and regulations, but since the scope of these laws and regulations is diverse and interrelated, the "Outline of Systematic Development for Automated Driving" was formulated in 2018 to clarify the government's overall policy toward the development of systems for AD. Under the close cooperation of the relevant ministries and agencies, the development of related laws and regulations is being promoted.

Specifically, the Road Transport Vehicle Act (Act No. 185 of 1951) and the Road Traffic Act (Act No. 105 of 1960) were revised and came into effect in April 2020, allowing SAE Level 3 AVs to drive on public roads.[15] In addition, the government has submitted a bill to amend the

[13] The Prime Minister's Official Residence Website. "Public-private ITS initiative and roadmap: Past efforts and basic concept of future ITS initiative." Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society and Public-Private Data Utilization Promotion Strategy Council, 15 Jun. 2021, https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20210615/roadmap.pdf
[14] Ibid.
[15] Ibid.

Road Traffic Law to the ordinary Diet session in 2022 to establish a permit system for the use of SAE Level 4 AVs, which fully automate driving under certain conditions, in local transportation services.[16] This bill was approved by the same session of the Diet.[17]

The "Guideline for Safety Technologies for Automated Vehicles" was formulated based on the fact that the "Outline of System Development for Automated Driving" calls for the compilation of guidelines by the summer of 2018 on safety requirements and safety assurance measures to be met by AVs. In the guideline, the definition of vehicle safety that must be met by AVs is "the absence of unacceptable risks"; i.e., the absence of reasonably foreseeable and preventable accidents involving bodily injury caused by ADS in the ODD of AVs.[18] In addition, the Announcement that Prescribes Details of Safety Regulations for Road Vehicles (Ministry of Land, Infrastructure, Transport and Tourism Announcement No. 619 of 2002) was revised in response to the revision of the Road Vehicles Act, and it stipulates that, with regard to the functions and performance of AVs, if there is a risk of collision with other vehicles or obstacles while the ADS is in operation, the system shall be controlled to reduce damage in the event of a collision to the maximum extent possible.[19] Although the provisions of the guideline and announcement can be understood as assuming a dilemma situation in the trolley problem, it is not always clear from them what is meant by "reasonably foreseeable and preventable accidents"

[16] Nikkei. "The Japanese government has approved a draft law to lift the ban on "Level 4" automated driving." 4 Mar. 2022, https://www.nikkei.com/article/DGXZQOUE034G70T00C22A3000000/

[17] Asahi. "Revised Road Traffic Law passed to allow fully automated driving under certain conditions." 19 Apr. 2022, https://www.asahi.com/articles/ASQ4M4K7NQ4LUTIL02F.html

[18] Automobile Bureau, The Ministry of Land, Infrastructure, Transport and Tourism. "the Guidelines for Safety Technologies for Automated Vehicles", https://www.mlit.go.jp/common/001253665.pdf.

[19] The Announcement that Prescribes Details of Safety Regulations for Road Vehicles Article 72-2, Item 7. If there is a risk of collision with other traffic or obstacles while the automated operation system or risk minimization control is in operation, or while the alarm of item 3 or 4 is issued, the control to prevent collision or to reduce damage in the event of collision to the maximum extent possible shall be activated.

and "control to reduce damage in the event of a collision to the maximum extent possible." Therefore, it is difficult to determine whether they assume the trolley problem or not.

Thus, at this point, there is no indication externally that the trolley problem related to AD is being explicitly discussed or considered by the Japanese government from an ethical perspective. This may be because, at this point, the ODD of AD is highways and the speed limit for AVs is low, making it difficult to foresee that a trolley problem dilemma will arise.

Government Policies and Initiatives for Automated Vehicles in the United States

In the United States, the federal government is responsible for regulating the safety performance of vehicles and vehicle equipment, as well as interstate commerce, while state and local governments play a leading role in licensing drivers, enacting traffic regulations, and developing policies on tort liability and insurance. These traditional roles are also considered to be the case for AVs.[20] For this reason, each state government has been working on its legislation regarding regulations, etc. for demonstration tests of AVs, while the federal government is also promoting various measures for the realization of AVs.

In the federal government, in May 2013, the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation (USDOT) released the "Preliminary Statement of Policy Concerning Automated Vehicles" as a preparatory document for policy on AVs. The document defines the levels of AD, outlines the agency's research plans, and provides recommendations to state governments on the use of experimental AVs on public

[20] The United State Department of Transportation. "Preparing for the future of transportation: Automated vehicles 3.0" 4 Oct. 2018, https://www.transportation.gov/av/3.

roads.[21] Subsequently, in September 2016, based on the progress of technological development related to AD, USDOT established the "Federal Automated Vehicles Policy," which aims to provide the initial regulatory framework and best practices to lead manufacturers and other organizations in the safe design, development, testing, and deployment of AVs.[22] The document addresses issues in facilitating the safe introduction and deployment of highly AVs, including vehicle performance guidance for AVs, state government model measures, and NHTSA regulatory measures for AVs. Note that the document is being issued as federal guidance rather than as a legally binding order, so as to quickly provide an initial regulatory framework and best practices to guide manufacturers and other involved parties.

In September 2017, the "Federal Automated Vehicles Policy" was replaced by the "Automated Driving System 2.0: A Vision for Safety," with the goal of supporting the review and design of best practices for the testing and deployment of AV technologies in the automotive industry, the United States, and other key stakeholders.[23] The document is intended to ensure a unified national framework for AD by clarifying the roles of the federal and state governments in regulating AVs and providing a framework that can be used by state governments in formulating laws and regulations.

In addition, in October 2018, the "Preparing for The Future of Transportation: Automated Vehicles 3.0," which outlines six principles to guide USDOT's policy on AVs and five implementation steps to put the principles into action, was developed.[24] The document requires

[21] The National Highway Traffic Safety Administration. "Preliminary statement of policy concerning automated vehicles." May. 2013, https://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf.
[22] The National Highway Traffic Safety Administration. "Federal automated vehicles policy." 20 Sep. 2016, https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016.
[23] The National Highway Traffic Safety Administration. "Automated driving system 2.0: A vision for safety." 12 Sep. 2017, https://www.nhtsa.gov/document/automated-driving-systems-20-voluntary-guidance.
[24] The United State Department of Transportation. "Preparing for the future of transportation: Automated vehicles 3.0." 4 Oct. 2018, https://www.transportation.gov/av/3.

developers of ADSs to conduct and publish a safety self-assessment of the 12 elements of AV safety outlined in the "Automated Driving System 2.0: A Vision for Safety." In addition, state governments are to be encouraged to remove unnecessary and incompatible regulatory and other barriers to AD technology and to support state interoperability.

Most recently, in January 2020, the "Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0" was published.[25] This document differs from previous ones. It describes the federal government's authority, research, and investment in AD to ensure a consistent approach by the federal government to AD technology and to ensure that the United States continues to lead the world in the research and development of AD technology. The document is intended to share information with industry, state governments, and other parties involved. While the previous documents basically focused on USDOT, this document covers 38 federal departments, agencies, commissions, and the Office of the President, demonstrating that the United States plans to continue playing a leadership role in the world in new technologies, including AVs. It is a statement of the U.S.' commitment to sustained global leadership in AVS and other new technologies.

Of these documents, the one that mentions the trolley problem is the "Federal Automated Vehicles Policy." In order to ensure the reasonable safety of ADSs, this document calls for the consideration of the following guidance items by all companies planning to manufacture, design, test, or sell ADSs in the United States: (1) data recording and sharing; (2) privacy; (3) system safety; (4) vehicle cybersecurity; (5) human-machine interface; (6) crashworthiness; (7) consumer education and training; (8) registration and certification; (9) post-crash behavior; (10)

[25] The National Science & Technology Council and the United State Department of Transportation. "Ensuring American leadership in automated vehicle technologies: Automated vehicles 4.0." 8 Jan 2020, https://www.transportation.gov/av/4.

federal, state, and local laws; (11) ethical considerations; (12) operational design domain; (13) object and event detection and response; (14) fallback (minimal risk conditions), and (15) validation methods.[26] In (11) ethical considerations, the programming of ADSs to address the dilemma regarding trolley problems should be widely accepted because it may affect not only AVs and their occupants, but also users in the surrounding roadways. The document also states that the programming of ADSs should be developed transparently with input from federal and state regulatory authorities, drivers, passengers, and vulnerable road users.[27] On the other hand, the revised version of the document, the "Automated Driving System 2.0: A Vision for Safety," states that the following guidance items should be considered by companies: (1) system safety; (2) operational design domain; (3) object and event detection and response; (4) fallback (minimal risk conditions); (5) validation methods; (6) human-machine interface; (7) vehicle cybersecurity; (8) crashworthiness; (9) post-crash behavior; (10) data recording; (11) consumer education and training; and (12) federal, state, and local laws. Ethical considerations are missing from the list.[28]

It is not clear from the document why "ethical considerations" were dropped from the guidance, but it is stated at the end of the document that the "NHTSA acknowledges that Privacy and Ethical Considerations are also important elements for entities to deliberate."[29] Although this document provides non-legally binding guidance for companies to consider based on the current state of technological development of ADSs, it encourages companies to publish a self-assessment of the safety of ADSs based on the guidance. Therefore, one of the reasons why

[26] The National Highway Traffic Safety Administration. "Federal automated vehicles policy." 20 Sept. 2016, https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016.
[27] Ibid.
[28] The National Highway Traffic Safety Administration. "Automated driving system 2.0: A vision for safety." 12 Sept. 2017.
[29] Ibid.

"ethical considerations" were dropped from the guidance items is that it is difficult to assume that a trolley problem dilemma will arise with regard to AVs at this time.

As mentioned above, each state government has been legislating its own regulations pertaining to AVs, and Congress is considering the enactment of federal laws as a unified rule for the United States since the requirements pertaining to regulations differ from state to state. In September 2017, the "SELF-DRIVE ACT" was passed by the House of Representatives.[30] This act stipulates the review of the "Federal Motor Vehicle Safety Standards" to ensure the safety of AVs, the requirement for companies to submit safety assessment certifications for AVs, and the establishment of the "Highly Automated Vehicle Advisory Council." The Senate debated the "AV START ACT," a bill that would make changes to the "SELF-DRIVE ACT,"[31] but it has yet to be passed due to a fatal accident caused by an Uber AV in March 2018.[32, 33] These Acts include cybersecurity vulnerabilities, encouragement of information disclosure, and concrete measures to protect privacy. These Acts do not include any provisions directly related to the trolley problem.

Under these circumstances, the "SELF-DRIVE ACT" was submitted to the House of Representatives again in September 2020, but there has been some criticism about the content of

[30] Congress.gov. "H.R.3388 - SELF DRIVE Act 115th Congress (2017-2018)" https://www.congress.gov/bill/115th-congress/house-bill/3388/actions?q=%7B%22search%22%3A%5B%22Safely+Ensuring+Lives+Future+Deployment+and+Research+In+Vehicle+Evolution+Act%22%2C%22Safely%22%2C%22Ensuring%22%2C%22Lives%22%2C%22Future%22%2C%22Deployment%22%2C%22and%22%2C%22Research%22%2C%22In%22%2C%22Vehicle%22%2C%22Evolution%22%2C%22Act%22%5D%7D&r=4&s=6.

[31] Congress.gov. "S.1885 - AV START Act 115th Congress (2017-2018)" https://www.congress.gov/bill/115th-congress/senate-bill/1885/all-actions?q=%7B%22search%22%3A%5B%22AV+START+Act%22%2C%22AV%22%2C%22START%22%2C%22Act%22%5D%7D&r=8&overview=closed&s=4#tabs.

[32] McFarland, Matt. "Uber self-driving car kills pedestrian in first fatal autonomous crash" CNN, 19 Mar. 2018, https://money.cnn.com/2018/03/19/technology/uber-autonomous-car-fatal-crash/index.html?iid=EL.

[33] Congress.gov. "H.R.3711 - SELF DRIVE Act 117th Congress (2021-2022)" https://www.congress.gov/bill/117th-congress/house-bill/3711/all-actions?s=3&r=1.

the Act, which mandates the establishment of uniform safety standards at the current stage of the development of AVs.[34]

Discussion in the United States

The Trolley Problem

Before taking a general look at the trolley problem with regard to AVs, I will describe this thought experiment in ethics. Originally conceived by the English philosopher Philippa Foot and formulated by the American philosopher Judith Jarvis Thomson, the trolley problem is a thought experiment with a wide range of cases.[35] The following two typical examples are often cited in contrast to one another.[36, 37]

> Case 1: A trolley is running out of control. There are five workers at the end of the track, and they will be run over and killed by the trolley. You are near the track and notice a lever nearby that pulls the trolley onto a branch line. There is another worker on the branch line, and, if you switch tracks, he will be run over and killed by the trolley, but you will be able to save the five workers. Should you pull the lever and switch the track or not?

> Case 2: A trolley is running out of control. There are five workers at the end of the track, and they will be run over and killed by the trolley. You are on a bridge over the track and notice a fat man peering over the track on the same bridge. If you push the man onto the track, the trolley will stop and the five workers will be saved, but the man will be killed by the trolley. Since you are small, you are unlikely to be able to stop the trolley even if you jump off onto the track. Should you push the fat man off or not?

[34] Advocates for Highway and Auto Safety. "Statement in opposition to the Self Drive Act." 23 Sept. 2020, https://saferoads.org/2020/09/23/self-drive-act/.
[35] Smith, Brent. "Personality facets and ethics positions as directives for self-driving vehicles." Technology in Society, 57, 2019, p.115–24.
[36] Foot, Philippa. "The problem of abortion and the doctrine of the double effect." Oxford Review, 1967.
[37] Thomson, Judith Jarvis. "The trolley problem." 1985.

From the viewpoint of utilitarianism, the right thing to do in Case 1 would be to pull the lever and push the man off in Case 2 to save the five people, because the action that results in the greatest happiness of the greatest number is considered right. On the other hand, from the viewpoint of Kantian deontology, one should not take advantage of others to realize one's goal, so one would not pull the lever in Case 1 and would not push the man off in Case 2.

1. Philippa Foot and the Trolley Problem

Philippa Foot, in a paper published in 1967,[38] addressed the doctrine of the double effect in relation to the abortion issue. In her paper, several ethical dilemmas were presented, one of which is the origin of the trolley problem. The doctrine of the double effect is a principle in ethics, originating in Catholic moral theology, that distinguishes between the intended effects of an action and the effects that were foreseen but not desired, and indicates under what circumstances it is permissible to cause the latter effect. Specifically, the following four conditions must be met: (1) the act itself must not be bad, (2) a good result must not be obtained by bad means, (3) the bad result must be foreseen, but not intended, and (4) there must be an important enough reason to tolerate a bad result.[39, 40]

Foot presented the following two cases in the paper to explore the doctrine of the double effect.

> Case 3: A mob says, "Find the perpetrator of a certain crime. I will kill five hostages if the perpetrator is not found." Since the perpetrator is unknown, the judge can only save the hostages by framing an innocent person for the crime and sentencing him to death. What should the judge do?

[38] Foot, Philippa. "The problem of abortion and the doctrine of the double effect." Oxford Review, 1967.
[39] Aulisio, Mark P. "Double effect, principle or doctrine of." Encyclopedia of Bioethics. 3rd edition, pp. 685-90.
[40] Sato, Hideaki. "Automated vehicles and the trolley problem." Chuo Gakuin University Human and Nature Series, 48, 2020, pp. 21-54.

> Case 4: A driver on a trolley that has lost control can only switch the track it is traveling on to another track. If he does not switch the track on which he is traveling, five workers on the track will be killed. There is one worker on another track, and switching tracks will sacrifice that one worker. What should the driver do?

With both cases, the question is whether it is acceptable to sacrifice one life to save five lives. After noting that sacrificing one life in Case 4 is considered acceptable, but not acceptable in Case 3, she explained the difference using the doctrine of the double effect. In Case 4, the driver intends to save five lives, but sacrificing one life was not intended, but only foreseen. On the other hand, in Case 3, the judge intends to save five lives and also to sacrifice one life. Therefore, she explained that, based on uses the doctrine of the double effect, sacrificing one person is acceptable in Case 4, but not in Case 3.

Foot then presented the following case and developed her critique of the doctrine of the double effect.

> Case 5: A hospital has five patients, and the production of a certain gas can save the five lives, but the gas inevitably releases a lethal gas into the room of another patient. If the other person cannot be moved from the room for some reason, should the gas be produced or not?

In this case, Foot argued that the act of producing a certain gas may be justified under the doctrine of the double effect, since the sacrifice of one patient's life was only foreseen, not intended, but that conclusion was considered unacceptable. Thus, she proposed the idea of a distinction between positive and negative obligations in place of the third principle of the doctrine of the double effect, the distinction between foresight and intention. This is based on whether an act falls under a positive or negative obligation, with priority given to the negative obligation in cases of conflict between positive and negative obligations, and comparative

balancing of outcomes in cases of conflict between obligations of the same kind. Based on this theory, in Case 5, since the positive obligation to save the lives of five patients conflicts with the negative obligation not to kill one patient, priority is given to the latter and the gas is not produced.[41]

2. Judith Jarvis Thomson and the Trolley Problem

In a paper published in 1985[42,43] Judith Jarvis Thomson pointed out that there were serious problems with Foot's paper by presenting Case 1 (the Bystander Case) about a trolley that has lost control.

Case 1 differs from Case 4 (the Driver Case) in that the driver is responsible for the safety of those who may be harmed by the trolley he is driving, while the bystander is just there by chance. If the driver does not switch tracks, he will run the trolley toward the five people and kill them. On the other hand, the bystander does not kill the five people, but leaves them to die, because the bystander himself does not run the trolley toward the five people without switching tracks. If the bystander switches tracks, however, he will run the trolley toward one person, which will kill him or her. Thus, the bystander's choice is to switch tracks and kill one person or not switch tracks and let five people die. Based on Foot's theory, it would be a conflict between negative and positive obligations, and bystanders should not switch tracks. Thomson found that conclusion puzzling and called it the "trolley problem," which she decided to examine.

---

[41] Philippa Foot, in the same paper, expresses her view on the three types of abortion issues based on the distinction between positive and negative obligations, which are not directly relevant to the trolley problem and are therefore omitted from this paper.
[42] Thomson, Judith Jarvis. "Killing, letting die, and the trolley problem." The Monist, 59.2, 1976, pp. 204-17.
[43] Thomson, Judith Jarvis. "The trolley problem." 1985.

Thomson presented Case 2 (the Fat Man Case) and examined why it was considered acceptable for a bystander to switch tracks in Case 1, while it was not considered acceptable for a bystander to push the fat man off the bridge in Case 2. As a result, if the means of saving five people must violate the rights of one person, as in Case 2, the bystander cannot choose that means. On the other hand, if the means of saving five people does not have to violate the rights of one person, as in Case 1, the bystander may choose that means.

In a paper published in 2008,[44] Thomson changed her position and argued that it was not permissible for bystanders to switch tracks in Case 1. Then she added a third option to Case 2 for consideration.

> Case 6: If the bystander turns the lever to the right, the trolley will enter the right-hand track, which will kill one worker. If the lever is turned to the left, the trolley will enter the left-hand track, which will kill the bystander himself, since he is standing on the track. If he does not operate the lever, he will leave the five workers to die. The bystander's options are to: (1) do nothing and let the five dies, (2) flip the switch to the right and kill one person, or (3) flip the switch to the left and kill himself.

Thomson argued that it was not permissible for a bystander to turn the trolley toward one worker because he was making that one worker pay for his good deed (saving five lives).[45] Since the only means left to do a good deed (saving five lives) is for the bystander to choose to kill himself, and it is difficult to make the bystander do that, it is acceptable not to do that good deed, i.e., it is acceptable to let the five people die.

Now, if Case 1 is examined again, there is no option to kill oneself, so the only way to save five lives is to sacrifice one. Thomson stated that, even though it was acceptable to let five

[44] Thomson, Judith Jarvis. "Turning the trolley." Philosophy & Public Affairs, 36.4, 2008, pp. 359-74.
[45] Judith Jarvis Thomson asserted that sacrificing one's own life when one could save five others by dying oneself was a sign of moral deficiency as a person.

people die when there were three options, it was not acceptable to make one person pay the compensation and save five lives when there were two options and that it was acceptable to let five people die in Case 1. Thomson also considered Case 4 by introducing a third option. Since it is the driver's responsibility to ensure the safety of the workers, Thomson asserted that, unlike Case 1, the driver could choose to sacrifice his own life. She did not state that when there were two options, as in Case 1, in which she said it was not acceptable to sacrifice one person to save five. This is because, according to Foot's theory (the distinction between negative and positive obligations), if the driver does nothing, he kills five people, whereas if the bystander does nothing, he lets five people die.

Thus, based on Foot's theory, in the conflict between the negative obligation (not to kill one person) and the positive obligation (to save five people) in Case 1, the former takes precedence and bystanders should not switch tracks. But Thomson found that conclusion puzzling and called it the trolley problem, which she examined. Later, she changed her argument and held that it was permissible for the former to take precedence, thus resolving the trolley problem in her 1985 paper. In this narrow sense, the trolley problem has been resolved, but this is only with respect to discussions within the scope discussed by Thomson. Foot and Thomson both rely on utilitarianism, so the trolley problem continues to be unresolved. The term the trolley problem is now used more broadly as a thought experiment in ethics, as described above, to ascertain whether it is permissible to sacrifice several people to help one.

The Trolley Problem with Regard to Automated Vehicles

The question of how AVs should be programmed to respond in advance when faced with a life-or-death dilemma for passengers and pedestrians has been the subject of research in the

United States and around the world.[46,47] In studies of such dilemmas, scenarios in which the trolley in the trolley problem is replaced by an AV are being put forward. Although some criticize the use of algorithms for AVs based on the trolley problem, there are high expectations for this type of research in terms of revealing what consumers think about AVs and, in turn, contributing to the widespread use of AVs.[48]

In this section, I would like to take a general view of the debate in the United States on the trolley problem as it relates to AVs. In the United States, many studies have been conducted, including statistical studies by Jean-Francois Bonnefon and his co-authors To compare this with the situation regarding the debate in Japan, I will present the results of several statistical studies of algorithms for AVs, mainly in terms of the suitability of installing utilitarian algorithms for

[46] Statistical studies other than those discussed in this paper include the following.

① Frison, Anna-Katharina, and Philipp Wintersberger, Andreas Riener. "First-person trolley problem: Evaluation of drivers' ethical decisions in a driving simulator." Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications, 2016, pp. 117-22.

② Sütfeld, Leon R., and Richard Gast, Peter König, Gordon Pipa. "Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure." Frontiers in behavioral neuroscience 11:122, 2017.

③ Frank, Darius-Aurel, and Polymeros Chrysochou, Panagiotis Mitkidis, Dan Ariely. "Human decision-making biases in the moral dilemmas of autonomous vehicles." Scientific Reports, vol. 9, no. 1, NATURE PUBLISHING GROUP, 2019, pp. 13080–119.

④ Kallioinen, Noa, and Maria Pershina, Jannik Zeiser, Farbod Nosrat Nezami, Gordon Pipa, Achim Stephan, Peter König. "Moral judgements on the actions of self-driving cars and human drivers in dilemma situations from different perspectives." Frontiers in Psychology, vol. 10, Frontiers Research Foundation, 2019, p. 2415.

⑤ Zhu, Anrun, and Shuangqing Yang, Yunjiao Chen, Cai Xing. "A moral decision-making study of autonomous vehicles: Expertise predicts a preference for algorithms in dilemmas." Personality and Individual Differences 186, 2022, p. 111356.

[47] Statistical studies outside the United States include the following.

（Germany）Faulhaber, Anja K., and Anke Dittmer, Felix Blind, Maximilian A. Wächter, Silja Timm, Leon R. Sütfeld, Achim Stephan, Gordon Pipa, Peter König. "Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles." Science and engineering ethics, 25.2, 2018, pp. 399-418.

（Republic of Finland）Radun, Igor, and Jenni Radun, Jyrki Kaistinen, Jake Olivier, Göran Kecklund, Tòres P. G. Theorell. "Endangering yourself to save another: A real-life ethical dilemma." Transportation Research. Part F, Traffic Psychology and Behaviour, 64, 2019, pp. 318–22.

[48] Novak, Thomas P. "A generalized framework for moral dilemmas involving autonomous vehicles: A commentary on Gill." The Journal of Consumer Research, 47(2), 2020, pp. 292–300.Other papers that mention trends in consideration of algorithms for AVs in the case of a socially agreed upon moral code include the following. Wiseman, Yair, and Ilan Grinberg. "The trolley problem version of autonomous vehicles." The Open Transportation Journal 12 (1), 2018, pp. 105–13.

dilemma situations.[49] The overall picture of statistical studies in the United States is described in detail in a 2020 paper by Thomas P. Novak.[50]

1. Statistical Studies Utilizing the Trolley Problem

a. Research by Jean-Francois Bonnefon et al.

Jean-Francois Bonnefon and his co-authors were probably the first to conduct systematic statistical study on the trolley problem as related to AVs in the United States. They said that creating an algorithm for AVs to deal with dilemma situations is a very difficult task as it must satisfy three incompatible factors: there must be consistency, it cannot be a sensation in the world, and it cannot discourage consumers from purchasing AVs.[51] They conducted six surveys (total of 1,928 participants) between June and November 2015.[52,53]

In Study 1, when asked whether an AV should sacrifice one passenger or 10 pedestrians in a dilemma situation, 76 percent of the participants said it should sacrifice one passenger.

[49] The following statistical studies on the trolley problem involving AVs in Japan all show that participants consider utilitarian algorithms morally desirable, but are less willing to purchase AVs equipped with such algorithms. Morita Tamayuki, and Shunsuke Managi. "Demand and social dilemmas created by automated driving cars," Research Institute of Economy, Trade and Industry, 2018.
Kawashima, Shigeo, and Satoshi Kitamura, Yasufumi Shibanai. "Attitudes toward the 'trolley problem' and other issues of automated vehicles: through a questionnaire survey of people living in Japan." Presentation Material for the 2017 Annual Conference of the Japan Society for Socio-Information Studies, 2017.

[50] Thomas P. Novak categorized recent statistical studies using scenarios based on the trolley problem using three factors: 1) participants' perspectives, 2) characters in the scenario, and 3) the nature of the vehicles in the scenario. He then classified four patterns in 1): (1) the perspective of a passenger in an AV, (2) the perspective of a pedestrian, (3) the perspective of an outside observer, and (4) the perspective of an anthropomorphic AV; in 2): (1) passengers are participants themselves and pedestrians are others, (2) passengers are others and pedestrians are participants themselves, and (3) both passengers and pedestrians are others; in 3): (1) manually operated vehicles, (2) AVs not operated by humans, and (3) AVs operated by humans only in emergency situations. Novak noted that future research should be conducted on combinations of these patterns for which no prior statistical study exists. Thomas P. Novak, "A Generalized Framework for Moral Dilemmas Involving Autonomous Vehicles: A Commentary on Gill." The Journal of Consumer Research, 47(2), 2020, pp. 292–300.

[51] Bonnefon, Jean-Francois, and Azim Shariff, Iyad Rahwan. "Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars?" 2015.

[52] Ibid.

[53] Bonnefon, Jean-Francois, and Azim Shariff, Iyad Rahwan. "The social dilemma of autonomous vehicles." Science (American Association for the Advancement of Science) 352 (6293), 2016, pp. 1573–76.

When asked to rate the moral algorithm of an AV on a scale from 0 (protect passengers) to 100 (minimize casualties), an overwhelming moral preference was shown for the utilitarian algorithm (median 85). In Study 2, when the number of pedestrians was varied from 1 to 100, the participants believed that passengers should not be sacrificed when there was only one pedestrian (mean approval rate 23 percent), but as the number of pedestrians increased, the approval rate also increased. In Study 3, participants were asked to indicate their likelihood of purchasing an AV with a utilitarian algorithm and an AV that prioritizes passenger protection on a scale of 1 to 100, respectively; the likelihood of purchasing an AV that prioritizes passenger protection was not high (median 50), but the participants were less likely to purchase an AV with a utilitarian algorithm (median 19). In Study 4, for each of the three algorithms of self-driving cars that (1) sacrifice one pedestrian to save ten pedestrians, (2) sacrifice one passenger to save ten pedestrians, and (3) sacrifice one other pedestrian to save one pedestrian, the participants were asked to answer 1) whether the algorithm is moral, 2) whether they would be willing to have an AV owned by another person carry the algorithm, and 3) whether they would purchase an AV with the algorithm. (1) received high ratings and (3) received low ratings, while (2) was rated high**,** though the willingness to purchase such an AV was low. In Study 5, when respondents were asked about the acceptability of legally enforcing a utilitarian algorithm on AVs, the acceptability of doing so was higher than that of enforcing it on human drivers. In Study 6, when respondents were asked to indicate their likelihood of considering purchasing an AV if the government regulated an algorithm for the vehicle and mandated a utilitarian algorithm, the median value without regulations was 59, while the median value with regulations was 21.

The studies revealed that people morally value an AV with a utilitarian algorithm and would like others to purchase one, while they would prefer to drive an AV with an algorithm that

prioritizes passenger protection. People were also found to be reluctant to allow government regulation of algorithms for AVs. Bonnefon and his co-authors pointed out that the results show typical features of social dilemmas in which people agree on what to do for the greater good, but do not take such actions themselves. They feared that, if regulations mandating utilitarian algorithms for AVs were introduced, many people would oppose them, which would delay the widespread use of AVs, and that many lives that would be saved by the widespread use of AVs would be lost. On the other hand, as long as people have a moral evaluation of AVs equipped with a utilitarian algorithm, a social norm may eventually be formed that strongly supports its adoption.

b. Research by Edmond Awad et al.

In a paper published in 2018,[54] Edmond Awad and his co-authors argued that it is not appropriate to leave devising algorithms for AVs in dilemma situations solely to either the technologists or ethicists and that it is necessary to clarify how such an algorithm was devised in order to encourage users to switch from human-driven vehicles to AVs and for the general public to accept the widespread use of AVs. They then conducted an extensive study of people around the world utilizing several dilemma situations to gauge societal expectations on the question of how AVs should respond to scenarios such as the trolley problem.

After collecting 40 million responses from people in 233 countries and regions, the following trends were identified. In dilemma situations, respondents were asked to answer whether AVs should sacrifice (1) humans or animals, (2) many or one person, (3) children or the elderly, (4) traffic law abiders or violators, (5) people of high social status or others, (6)

[54] Awad, Edmond, and Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-Francois Bonnefon, Iyad Rahwan. “The moral machine experiment.” Nature (London) 563 (7729), 2018, pp. 59–64.

physically unfit or fit people, (7) women or men, or (8) passengers or pedestrians. The tendency to sacrifice the latter over the former was observed, and this tendency was even stronger especially for (1) through (3). There were also differences in responses by region and culture, with participants largely divided into (1) the Western cluster (North America and many Protestant, Catholic, and Orthodox Christian European countries), (2) the Eastern Cluster (Far Eastern countries such as Japan and Taiwan, and Islamic countries such as Indonesia, Pakistan, and Saudi Arabia), and (3) the Southern Cluster (Latin American countries, etc.). For example, the tendency to sacrifice the elderly rather than children was less pronounced in the Eastern Cluster and much stronger in the Southern Cluster, as was the tendency to sacrifice those of lower social status rather than those of higher social status.[55]

Based on these results, Awad and his co-authors found, for example, that if policymakers decide that the algorithm for AVs should sacrifice children rather than the elderly, they face the challenge of explaining the rationale for such a decision and the inevitable strong backlash if children are actually killed in an AV traffic accident.[56] On the other hand, they stated that, although it was not possible to find universal agreement in their research, the fact that relatively consensual preferences were found in a wide range of regions around the world indicates that the path to devise a consensus-based algorithm for AVs is not hopeless from the outset.

[55] In the same paper, Edmond Awad et al. wrote, "our data reveal a set of preferences in which certain characters are preferred for demographic reasons. First, we observe that higher country-level economic inequality (as indexed by the country's Gini coefficient) corresponds to how unequally characters of different social status are treated. Those from countries with less economic equality between the rich and poor also treat the rich and poor less equally in the Moral Machine. This relationship may be explained by regular encounters with inequality seeping into people's moral preferences, or perhaps because broader egalitarian norms affect both how much inequality a country is willing to tolerate at the societal level, and how much inequality participants endorse in their Moral Machine judgments."

[56] The following paper made similar points.
Maxmen, Amy. "A moral map for AI cars." Nature, 562.7728, 2018, pp. 469–70.

c. Research by Yochanan E. Bigman et al.

Yochanan E. Bigman and his co-author criticized the conclusion by Awal and his co-authors that people tend to decide who to sacrifice based on personal characteristics (gender, age, status, etc.), saying that such a conclusion is contrary to the U.S. Constitution, the UN Universal Declaration of Human Rights, etc. Therefore, a third option, that "AVs should decide who to save and who to kill without considering personal characteristics," was introduced in their research, and responses were obtained from a total of more than 5,000 people in the United States and the U.K.[57]

In Study 1, respondents were asked to answer whether AVs should sacrifice one group (children, women, people of high social status, traffic law abiders, etc.) or another (elderly people, men, people of low social status, traffic law violators, etc.), respectively, and the research by Edmond Awad and his co-authors. Similar results were obtained, with 70 percent to 90 percent of participants sacrificing the latter over the former, respectively. In Study 2, when a third option of treating the former and the latter equally was introduced in Study 1, between 70 percent and 90 percent of the participants chose the former and the latter equally, respectively. In Study 3, when participants were asked to indicate whether they preferred an AV that made decisions about who to sacrifice based on structural characteristics (number of victims, whether they were passengers or pedestrians, etc.) or an AV that made decisions about who to sacrifice based on both structural and personal characteristics, 89.9 percent chose an AV that made decisions about who to sacrifice based on structural characteristics.

Bigman and his co-author stated that the findings that people want AVs to disregard personal characteristics and treat people equally are not only consistent with the U.S.

[57] Bigman, Yochanan E., and Kurt Gray. "Life and death decisions of autonomous vehicles." Nature, 579.7797, 2020, pp. E1–E2.

Constitution and other laws, but also with the current level of technology, which makes it difficult for AVs to discern the personal characteristics of pedestrians and other people.[58]

d. Research by Tripat Gill

Tripat Gill believed that a systematic examination of the reasons for the social dilemmas (people morally value AVs with utilitarian algorithms and want others to buy them, but they want to drive AVs with algorithms that prioritize passenger protection) identified in the research by Bonnefon and his co-authors was needed. Gill asserted that such a systematic examination would be necessary to ascertain: (1) the moral norms that people impose on themselves, (2) whether people's perspectives influence moral norms, and (3) how moral norms would be affected if the subject were changed from a human-driven vehicle to an AV, and conducted studies.[59]

In Study 1 (316 participants), participants were asked to respond to the question of whether a vehicle (human-driven or automated) should sacrifice one passenger or one pedestrian from the driver's (passenger) or pedestrian's perspective. As a result, from the driver's (passenger) perspective, 23 percent of the participants answered human-driven vehicles should sacrifice pedestrians and 52 percent of the participants answered AVs should sacrifice pedestrians. On the other hand, from the pedestrian's perspective, 27 percent of the participants answered human-driven vehicles should sacrifice pedestrians and 52 percent of the participants

[58] Nature, 579.7797, 2020, pp. E3–E5. In this paper, Edmond Awad et al. responded to the criticism of Yochanan E. Bigman et al. that the approach adopted by Bigman et al. makes participants more sensitive to social desirability and experimental demands by explicitly offering them the option of treating people equally and that the research by Awad et al. also allows for the response of treating people equally, since it asks for responses on a continuous scale with a sliding bar, rather than a simple choice of preferences based on personal characteristics. Awad, Edmond, and Sohan Dsouza1, Richard Kim1, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, Iyad Rahwan. "Reply to: Life and death decisions of autonomous vehicles."

[59] Gill, Tripat. "Blame it on the self-driving car: how autonomous vehicles can alter consumer morality." Journal of Consumer Research, 47.2, 2020, pp. 272-91.

answered AVs should sacrifice pedestrians. In Study 2 (269 participants), participants were asked to respond to the perception of the degree of responsibility for the driver's (passenger's) decisions in Study 1. The results revealed that participants rated their responsibility for victimizing pedestrians lower when they were in an AV compared to when they were driving a vehicle themselves. In Study 3 (284 participants), participants were asked to respond to the question limited to the driver's (passenger's) perspective in Study 1, with the number of pedestrians divided into two groups: one and five, respectively. The results showed that 20 percent of participants would sacrifice pedestrians with human-driven vehicles and 48 percent with AVs when there was just one pedestrian, 19 percent for human-driven vehicles, and 25 percent for AVs when there were five pedestrians. In Study 4 (349 participants), participants were asked to respond to the question limited to the driver's (passenger) perspective in Study 1, with pedestrians divided into adults and children. The results showed that 38 percent would sacrifice adult pedestrians with human-driven vehicles and 54 percent would do the same with AVs, while the corresponding figures were 9 percent and 15 percent of when the pedestrian was a child. In Study 5 (612 participants), participants were asked to respond to the question limited to the driver's (passenger's) perspective in Study 1, and the injuries sustained by the driver (passenger) or pedestrian were severe (including death) and moderate. As a result, the number of participants who answered that pedestrians should be sacrificed was 36 percent with human-driven vehicles and 53 percent with AVs when the injury was severe and 14 percent with human-driven vehicles and 26 percent with AVs when the injury was moderate.

Thus, the studies found that people were more likely to be willing to sacrifice a pedestrian if they were a passenger in an AV compared to if they were driving a vehicle themselves. This change in moral judgment was driven by a lower sense of responsibility for

traffic accidents caused by AVs, and the same trend was confirmed when the degree of damage was varied. This trend weakened, however, when there were five pedestrians or children. Based on these results, Gill stated that AVs have the potential to change the general ethical code and increase user self-interest.

2. Criticism of the Trolley Problem

As noted above, numerous studies have been conducted on the trolley problem with AVs, and there has also been a certain amount of criticism of them. Specifically, the following arguments have been presented.

(1) In the trolley problem, participants individually decide whether or not to steer an AV in an imminent situation, but in devising algorithms for AVs, multiple stakeholders weigh many considerations, and the decision-making process is different between the two.[60]

(2) In a real-life dilemma such as the trolley problem, not only how the AV will perform, but also how the other parties (pedestrians, oncoming vehicles, etc.) will react is an important factor, but it is difficult to predict the behavior of another party. In addition, the consequences of public roads for AVs are determined based on risk assessment based, in turn, on information about oneself and the other party, but in reality it is not always possible for AVs to obtain the information necessary for risk assessment (age, seat belt use, alcohol consumption, health condition, etc.) in an exhaustive manner. The existence of such uncertainty may hinder the algorithm-based actions of AVs.[61, 62, 63. 64, 65]

[60] Nyholm, Sven R., and Jillles Smids. "The ethics of accident-algorithms for self-driving cars: An Applied Trolley Problem?" Ethical Theory and Moral Practice 19 (5), 2016, pp. 1275–89.
[61] Ibid.
[62] Page, Misconceptions. "Top misconceptions of autonomous cars and self-driving vehicles." 2016.
[63] Smith, Bryant W. "The trolley and the pinto." Texas A&M Law Review 4 (2), 2017, pp. 197–208.

(3) Where the specific risk of harm depends on the probability and severity of the harm, the trolley problem has narrowed the discussion to only the severity of the harm to simplify the situation.[66, 67]

(4) Currently, it is extremely rare for a human-driven vehicle to face the dilemma assumed in the trolley problem, and the probability is expected to be even lower for AVs. Nevertheless, it is not practical to discuss the trolley problem, which may hinder the widespread use of AVs depending on the outcome of the discussion.[68, 69, 70, 71, 72]

---

[64] Lundgren, Björn. "Safety requirements vs. crashing ethically: What matters most for policies on autonomous vehicles." AI & Society, vol. 36, no. 2, Springer London, 2020, pp. 405–15.

[65] Mirnig, Alexander, and Alexander Meschtscherjakov. "Trolled by the trolley problem." Proceedings of the 2019 CHI conference on human factors in computing systems, ACM, 2019, pp. 1–10.

[66] Smith, Bryant W. "The trolley and the pinto." Texas A&M Law Review 4 (2), 2017, pp. 197–208.
Smith states that, unless the probability of harm or severity of harm on the road is zero, AVs must take both factors into account and make decisions based on the balance between them.

[67] In response to this criticism, Sebastian Krügel et al. stated that safety engineering principles are divided into primary prevention (reducing the probability of a hazard occurring) and secondary prevention (reducing the severity of a hazard), and that the engineering literature on AVs focuses on primary prevention. In addition, they found that, even when the concept of risk was introduced into the dilemma situation scenario, the tendency to sacrifice one person rather than five was seen.
Krügel, Sebastian, and Matthias Uhl. "Autonomous vehicles and moral judgments under risk." Transportation Research. Part A, Policy and Practice, vol. 155, Elsevier Ltd, 2022, pp. 1–10.

[68] Page, Misconceptions. "Top misconceptions of autonomous cars and self-driving vehicles." 2016.

[69] Johannes Himmelreich states that the trolley problem involving AVs is based on an inconsistent set of technical constraints because there is damage selection even though a collision is inevitable in the automated vehicle trolley problem scenario. Himmelreich, Johannes. "Never mind the trolley: The ethics of autonomous vehicles in mundane situations." Ethical Theory and Moral Practice, vol. 21, no. 3, Springer, 2018, pp. 669–84.

[70] It is worth noting that Jean-Francois Bonnefon et al, considered the pioneers in conducting statistical studies in the U.S., expressed this opinion. They then assumed a situation in which a large truck, an AV, and a bicycle are driving side-by-side, and argued that the decisions made by AVs in such situations are not as critical and directly life-or-death as the dilemma situations in the trolley problem, but they can occur normally and there is a small risk of accidents, and the risk may become apparent as a large number of such situations accumulate. They named this the "statistical trolley dilemma," and stated that AV engineers must solve this dilemma.
Bonnefon, Jean-Francois, and Azim Shariff, Iyad Rahwan. "The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view]." Proceedings of the IEEE, vol. 107, no. 3, IEEE, 2019, pp. 502–4.

[71] Lundgren, Björn. "Safety requirements vs. crashing ethically: What matters most for policies on autonomous vehicles." AI & Society, vol. 36, no. 2, Springer London, 2020, pp. 405–15.

[72] Freitas, Julian De, and Sam E. Anthony, Andrea Censi, George A. Alvarez1. "Doubting driverless dilemmas." Perspectives on Psychological Science, vol. 15, no. 5, SAGE Publications, 2020, pp. 1284–88.

(5) The question of how to devise algorithms for AVs to deal with dilemmas such as the trolley problem should be considered in light of the legal issue of who is liable in the event of a traffic accident caused by an AV, but this perspective is lacking in the trolley problem.[73, 74, 75]

(6) There is a big difference between considering what is right and what is wrong when examining ethical problems, and even though AVs are required not to make legally and ethically wrong decisions, AVs are required to make ethically correct decisions when it comes to the trolley problem.[76, 77]

(7) While previous research has shown a trend toward ethical judgments based on age, gender, and other factors, the Institute of Electrical and Electronics Engineers prohibits discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, and gender expression in its Code of Ethics.[78] The development of algorithms for AVs based on the results of the studies, then, would violate this code of ethics.[79]

---

[73] Nyholm, Sven R., and Jillles Smids. "The ethics of accident-algorithms for self-driving cars: An Applied Trolley Problem?" Ethical Theory and Moral Practice 19 (5), 2016, pp. 1275–89.

[74] Casey, Bryan. "Amoral machines, or: How roboticists can learn to stop worrying and love the law." Nw. UL Rev., 111, 2017, p. 1347..Casey states that legal liability for a company is a system of compensation for damages in an accident, and that a company's cost-benefit calculation minimizes liability (damages) rather than maximizing morality, and that a legal system that bridges the gap between moral norms and legal liability is required.

[75] Johannes Himmelreich argued that moral answers are pluralistic and are a matter of individual choice, while political answers are a social choice and a kind of compromise to overcome differences of opinion, and that in the trolley problem involving AVs, where a political solution is required, moral answers have been sought.
Himmelreich, Johannes. "Never mind the trolley: The ethics of autonomous vehicles in mundane situations." Ethical Theory and Moral Practice, vol. 21, no. 3, Springer, 2018, pp. 669–84.

[76] Page, Misconceptions. "Top misconceptions of autonomous cars and self-driving vehicles." 2016.

[77] Bradley, Wendel W. "Economic rationality and ethical values in design-defect analysis: The trolley problem and autonomous vehicles." California Western Law Review 55 (1), 2018, p. 129.
Bradley argues that whether a company is liable for the decisions of an AV in a dilemma situation depends on whether the victim (pedestrian) can prove that a reasonable alternative design (RAD) would have prevented the accident and that the AV was unreasonably dangerous by not implementing an alternative design. However, he stated that it would be difficult for a jury to find that a different algorithm (passenger death instead of pedestrian) is a RAD. He also stated that he hopes that the ADS would be able to intervene before the dilemma situation arises and avoid having to choose between one life or another.

[78] The Institute of Electrical and Electronics Engineers website. "IEEE code of ethics" Jun. 2016, https://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/corporate/ieee-code-of-ethics.pdf.

[79] Lin, Patrick. "Why ethics matters for autonomous cars. In: Autonomous driving." Springer, Berlin, Heidelberg, 2016, pp. 69-85. Lin pointed out that, in a dilemma situation, programming an AV to select collisions with certain targets over others is analogous to a targeting algorithm. Based on the statistical results that those wearing helmets

3. Reactions of Businesses to the Trolley Problem

a. Mercedes-Benz

In October 2016, Mercedes-Benz drew global attention when it became the first major automaker to offer an answer to the trolley problem involving AVs. Christoph von Hugo, the company's manager of driver assistance systems and active safety, stated that "[a]ll of Mercedes-Benz's future Level 4 and Level 5 autonomous cars will prioritize saving the people they carry" and that "[i]f you know you can save at least one person, at least save that one. Save the one in the car."[80, 81] His statement caused a media frenzy, but Mercedes-Benz said that he was misquoted and insisted that the company "remains committed to the principle of providing the highest level of safety for all road users."[82] In the same interview, Von Hugo also said,

> We believe that this ethical problem is not as relevant as people think today. It is a situation that cannot be addressed by current drivers, and from a physical point of view we cannot prevent it today, nor can AVs. It is just that (AVs) are much better than the average (human) driver. This is a moral question: who are we saving? Ninety-nine percent of our engineering work is to make sure that this situation does not happen at all. We work hard to make sure that AVs do not get into such situations, and that they stay away from potential situations where they would have to make such decisions.[83]

---

have a lower rate of serious injury in traffic accidents than those not wearing helmets, he argued that AVs could be programmed to choose to collide with the former over the latter (A similar point is also made by Goodall, Noah J. "Away from trolley problems and toward risk management," Applied Artificial Intelligence, 30.8, 2016, pp. 810-21.).

[80] Taylor, Michael. "Self-driving mercedes-benzes will prioritize occupant safety over pedestrians." Car and Driver, 7 Oct, 2016. https://www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/. In the above article, Christoph von Hugo argued: "You could sacrifice the car. You could, but then the people you've saved initially, you don't know what happens to them after that in situations that are often very complex, so you save the ones you know you can save."

[81] Brown, Mike. "Mercedes's self-driving cars will kill pedestrians over drivers." Inverse, 14 Oct, 2016. https://www.inverse.com/article/22204-mercedes-benz-self-driving-cars-ai-ethics.

[82] Casey, Bryan. "Amoral machines, or: How roboticists can learn to stop worrying and love the law." Nw. UL Rev., 111, 2017, p. 1347.

[83] Hern, Alex. "Self-driving cars don't care about your moral dilemmas." The Guardian, 22 Aug. 2016. https://www.theguardian.com/technology/2016/aug/22/self-driving-cars-moral-dilemmas.

b. Google-affiliated company

According to a 2016 article in *The Guardian,*[84] an engineer at Google's affiliate company X, which is leading the development of AVs, does not think that the trolley problem involving AVs is all that interesting and that they have never faced such a problem. He said that, if an AV were to encounter a situation where it had to choose between an infant and an elderly person, it would usually mean that the AV had made a mistake seconds before and that it was important to avoid such a situation altogether to save lives. He also noted that, when faced with such a situation, the answer is almost always "apply the brakes," and that AVs are more accurately controlled by stepping on the brakes than by steering them and driving into something.[85]

c. May Mobility

According to a presentation by Edwin Olson, CEO of May Mobility, a U.S. startup company developing automated shuttles,[86] the real world is so complex and the consequences of actions so uncertain that it is not rational to operate AVs based on moral preferences in the findings on the trolley problem with AVs. He then stated that a good choice in an uncertain environment is to brake to delay causing damage. He stated that allowing as much time as possible for each traffic entity to take crisis-avoidance action by allowing as much time as possible before damage occurs, as well as reducing the kinetic energy of the AV, can reduce the

[84] Bradley, Wendel W. "Economic rationality and ethical values in design-defect analysis: The trolley problem and autonomous vehicles." California Western Law Review 55 (1), 2018, p. 129.

[85] Davnall, Rebecca. "Solving the single-vehicle self-driving car trolley problem using risk theory and vehicle dynamics." Science and Engineering Ethics, vol. 26, no. 1, Springer Netherlands, 2019, pp. 431–49.
Davnall states that "In situations where an AV must choose between braking straight ahead for an unavoidable collision and turning for an unavoidable collision, the AV should always choose the straight ahead option" and "In the trolley problem of AVs, there are situations where the vehicle is within stopping distance and cannot avoid the collision by simply applying the brakes. Therefore, the steering maneuver performed by the AV will be abrupt compared to the speed, and there is a risk of losing control of the AV. In a crowded environment where a collision could occur, loss of vehicle control is far more dangerous than a controlled stop."

[86] Olson, Edwin. "Trolley folly." May Mobility, 14 Nov, 2018, https://medium.com/may-mobility/trolley-folly-fcbd181b7152.

damage in an accident. He also stated that the difference between a good driver and a bad one is not whether or not they are good at emergency maneuvers, but whether or not they are less likely to get into dangerous situations in the first place, and that the trolley problem with AVs distracts from that most important aspect of safe driving.[87]

Conclusion

This paper begins with an overview of government policies and initiatives for AVs in Japan and the United States. Both the U.S. and Japanese governments consider the realization and popularization of AVs as an important policy issue, and are promoting demonstration tests and examining legal issues. On the other hand, I could not necessarily confirm a situation in which the governments are actively discussing the trolley problem involving AVs from their particular standpoints. In the United States, a section on ethical considerations for AVs was included in a government document at one time, but this section was deleted from the revised document, indicating that a consistent stance on the trolley problem with AVs has not yet been determined. This is thought to reflect the current state of technologies and demonstration tests related to AVs. Since it is difficult to assume that the government will take the initiative in tackling the trolley problem involving AVs when the companies concerned are not actively considering this problem, the current studies have probably been mainly conducted from an academic standpoint. In any case, the U.S. and Japanese governments are likely to consider their

[87] Freitas, Julian De, and Sam E. Anthony, Andrea Censi, George A. Alvarez1. "Doubting driverless dilemmas." Perspectives on Psychological Science, vol. 15, no. 5, SAGE Publications, 2020, pp. 1284–88. De Freitas et al. state, "The four AVs companies we surveyed (May Mobility, autonomy, Perceptive Automata, and a global car company that wished to remain anonymous) have no team or budget to solve dilemma situations like the trolley problem, even though they put AVs on real roads every day." "Senior engineers and directors from the companies expressed the view that "teaching AVs to solve dilemma situations like the trolley problem is reckless and irrelevant to the safety of AVs. They rejected the idea of choosing who to kill based on a person's social category because of the practical, ethical, and legal issues involved."

stance on this problem while closely monitoring future technologies and the status of demonstration tests related to AVs.

In considering government stance, there is the issue of whether or not the government should regulate AV algorithms for dealing with dilemmas such as the trolley problem. Research by Bonnefon and his co-authors found that there are many objections to such regulation by the government,[88] and some have argued for delegating the choice of algorithms for AVs to their users.[89] Gill, however, argued that doing so would undermine cooperation and deepen conflict in society, and therefore, that the government should mandate algorithms to be installed in AVs based on a uniform code of ethics that is consistent with what human drivers practice and what pedestrians expect.[90] As mentioned above, however, the answer to the question of whom AVs will sacrifice in a dilemma such as the trolley problem depends on the viewpoint of the responding entities. Therefore, it is also argued that it is necessary to ensure that all parties involved (AV-related businesses, governments, consumers, etc.) cooperate in examining algorithms, and also consider how liability should be handled in the event of a traffic accident caused by an algorithm devised in this way, so that AV-related businesses do not bear excessive liability.[91, 92, 93]

---

[88] The results of a similar statistical study include the following. The Open Roboethics Initiative. "If death by autonomous car is unavoidable, who should die? Reader poll results." 23 Jun. 2014, https://robohub.org/if-a-death-by-an-autonomous-car-is-unavoidable-who-should-die-results-from-our-reader-poll/.

[89] Fournier, Tom. "Will my next car be a libertarian or a utilitarian?: Who will decide?" IEEE Technology and Society Magazine, 35.2, 2016, pp. 40-5. Giuseppe Contissa, Francesca Lagioia, Giovanni Sartor, "The ethical knob: ethically-customisable automated vehicles and the law." Artificial Intelligence and Law, 25.3, 2017, pp. 365-78.

[90] A similar claim is made in the following paper. Gogoll, Jan, and MÜLLER, Julian F Müller. "Autonomous cars: in favor of a mandatory ethics setting." Science and engineering ethics, 23.3, 2017, pp. 681-700.

[91] Gurney, Jeffrey K. "Crashing into the unknown: An examination of crash-optimization algorithms through the two lanes of ethics and law." Albany Law Review, vol. 79, no. 1, Albany Law School, 2015, p. 183.

[92] Wu, Stephen S. "Autonomous vehicles, trolley problems, and the law." Ethics and Information Technology, vol. 22, no. 1, Springer Netherlands, 2019, pp. 1–13.

[93] Krügel, Sebastian, and Matthias Uhl. "Autonomous vehicles and moral judgments under risk." Transportation Research. Part A, Policy and Practice, vol. 155, Elsevier Ltd, 2022, pp. 1–10.

Next, this paper reviews the results of statistical studies on the trolley problem involving AVs in the United States and elsewhere. Bonnefon and his co-authors found that people prefer AVs with utilitarian algorithms and want others to purchase them, but that they prefer AVs that prioritize passenger protection for themselves. Awad and his co-authors found that people not only select AVs with utilitarian algorithms, but also tend to consider the personal characteristics of the people involved in making decisions about who to sacrifice, with some regional and cultural variation in this tendency. Bigman and his co-author noted that, unlike the findings of Awad and his co-authors, people essentially wanted to treat everyone involved equally without regard to personal characteristics. Gill asserted that people are less likely to attribute responsibility for traffic accidents that occur to themselves when they are passengers in an AV compared to when they are driving a vehicle.

As for the consideration of personal characteristics in situations such as the trolley problem, attention should be paid to the "Ethical rules for automated and connected vehicular traffic"[94] published by the Ethics Committee of the German Federal Ministry of Transport and Digital Infrastructure in 2017. This committee was established in 2016 by the Federal Minister of Transport and Digital Infrastructure, and its members include individuals from the fields of philosophy, law, social sciences, technology impact assessment, the automotive industry, and software development. The ethical rules consist of a total of 20 articles, of which the ninth prohibits distinguishing between victims based on personal characteristics or endangering non-parties, although it is permissible to reduce the number of human victims in dilemmas such as

[94] The Federal Ministry of Transport and Digital Infrastructure. "Ethics commission automated and connected driving report 2017" 2017 https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile.

the trolley problem.[95] In other words, Germany explicitly prohibits AVs from taking personal characteristics into account when making decisions in dilemmas. The results of Bigman and his co-author are consistent with the ethical rules and will be helpful for future consideration by the U.S. and Japanese governments.

Research in the United States and other countries has identified a social dilemma between AVs with utilitarian algorithms and AVs that prioritize passenger protection, and the formation of this dilemma is influenced by the fact that people are less responsible for traffic accidents that occur when they are in AVs than when they are driving vehicles themselves. So, in examining the trolley problem involving AVs, one issue to be discussed is how to overcome this problem. Therefore, I would like to introduce one view that may help in examining this issue.[96]

Based on Thomson's discussion, if the trolley problem involving AVs is made to correspond to the Bystander Case, it would not be acceptable for AVs to sacrifice one pedestrian, even if the AVs are to save five pedestrians, on the assumption that the priority is protection of the AVs. On the other hand, if the trolley problem involving AVs is made to correspond to the Driver Case, it would be acceptable for AVs to sacrifice one pedestrian to save five pedestrians, on the assumption that AVs sacrifice themselves. Now, should we address the Bystander Case or the Driver Case for Level 4 and above AVs? Since passengers in Level 4 and above AVs are not involved in driving, this would create a situation in which the driver of an AV and the passengers of the AV (the persons who would suffer if the AV were to be victimized) are different. If the passengers of the AV must minimize the damage caused by the AV in a traffic accident, then the

[95] Kallioinen, Noa, and Maria Pershina, Jannik Zeiser, Farbod Nosrat Nezami, Gordon Pipa, Achim Stephan, Peter König. "Moral judgements on the actions of self-driving cars and human drivers in dilemma situations from different perspectives." Frontiers in Psychology, vol. 10, Frontiers Research Foundation, 2019, p. 2415.

[96] Sato, Hideaki. "Automated vehicles and the trolley problem." Chuo Gakuin University Human and Nature Series, 48, 2020, pp. 21-54.

Driver Case should be addressed, and, if they do not have such an obligation, then the Bystander Case should be addressed. On the other hand, the driver entity of an AV is considered to have an obligation to minimize the damage from a traffic accident caused by the AV, but it is the passengers of the AV, not the driver entity of the AV, who suffer when the AV is damaged. Therefore, it is considered undesirable if the passengers of an AV are not obligated to minimize the damage caused by an AV in a traffic accident, because the driver-subject of an AV damaging the AV in a traffic accident would impose a burden on the passengers of the AV to avoid the sacrifice of pedestrians. Therefore, it is necessary to consider whether the Bystander Case or the Driver Case should be addressed for Level 4 and above AVs, based on the responsibility of AV passengers in traffic accidents. Thus, it is considered necessary to consider the algorithm for AVs in dilemmas such as the trolley problem based on the nature of liability in traffic accidents involving AVs.

Although there are many critical opinions related to the trolley problem involving AVs as described in this paper, it is believed that there is some possibility that this will become a realistic problem that needs to be discussed depending on future technological developments related to AVs, etc. Even if the best option in dilemmas such as the trolley problem is to apply the brakes, as AVs developers claim, the "statistical trolley dilemma"[97] mentioned by Bonnefon and others will continue to be an issue for the realization of AVs. It is assumed that this will be one of the issues to be discussed with regard to the widespread use of AVs. I hope this paper will be of some help in such discussions.

[97] See footnote 69.

BIBLIOGRAPHY

Advocates for Highway and Auto Safety. "Statement in opposition to the SELF DRIVE ACT." 23 Sep. 2020, https://saferoads.org/2020/09/23/self-drive-act/

Aulisio, Mark P. "Double effect, principle or doctrine of." Encyclopedia of Bioethics. 3rd edition, pp. 685-690.

Automobile Bureau, The Ministry of Land, Infrastructure, Transport and Tourism. "the guidelines for safety technologies for automated vehicles." https://www.mlit.go.jp/common/001253665.pdf

Awad, Edmond, and Sohan Dsouza1, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, Iyad Rahwan. "Reply to: Life and Death Decisions of Autonomous Vehicles." Nature, 579.7797, 2020, pp. E3–E5.

Awad, Edmond, and Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-Francois Bonnefon, Iyad Rahwan. "The Moral Machine Experiment." Nature (London) 563 (7729), 2018, pp. 59–64.

Bigman, Yochanan E., and Kurt Gray. "Life and Death Decisions of Autonomous Vehicles." Nature, 579.7797, 2020, pp. E1–E2.

Bonnefon, Jean-Francois, and Azim Shariff, Iyad Rahwan. "Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?" 2015.

Bonnefon, Jean-Francois, and Azim Shariff, Iyad Rahwan. "The Social Dilemma of Autonomous Vehicles." Science (American Association for the Advancement of Science) 352 (6293), 2016, pp. 1573–1576.

Bonnefon, Jean-Francois, and Azim Shariff, Iyad Rahwan. "The Trolley, The Bull Bar, and Why Engineers Should Care About The Ethics of Autonomous Cars [point of View]." Proceedings of the IEEE, vol. 107, no. 3, IEEE, 2019, pp. 502–504.

Bradley, Wendel W. "Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles." California Western Law Review 55 (1), 2018, p. 129.

Brown, Mike. "Mercedes's Self-Driving Cars Will Kill Pedestrians Over Drivers." Inverse, 14 Oct, 2016. https://www.inverse.com/article/22204-mercedes-benz-self-driving-cars-ai-ethics

Casey, Bryan. "Amoral machines, or: How roboticists can learn to stop worrying and love the law." Nw. UL Rev., 111, 2017, p. 1347.

Congress.gov. “H.R.3388 - SELF DRIVE Act 115th Congress (2017-2018)” https://www.congress.gov/bill/115th-congress/house-bill/3388/actions?q=%7B%22search%22%3A%5B%22Safely+Ensuring+Lives+Future+Deployment+and+Research+In+Vehicle+Evolution+Act%22%2C%22Safely%22%2C%22Ensuring%22%2C%22Lives%22%2C%22Future%22%2C%22Deployment%22%2C%22and%22%2C%22Research%22%2C%22In%22%2C%22Vehicle%22%2C%22Evolution%22%2C%22Act%22%5D%7D&r=4&s=6

Congress.gov. “H.R.3711 - SELF DRIVE Act 117th Congress (2021-2022)” https://www.congress.gov/bill/117th-congress/house-bill/3711/all-actions?s=3&r=1

Congress.gov. “S.1885 - AV START Act 115th Congress (2017-2018)” https://www.congress.gov/bill/115th-congress/senate-bill/1885/all-actions?q=%7B%22search%22%3A%5B%22AV+START+Act%22%2C%22AV%22%2C%22START%22%2C%22Act%22%5D%7D&r=8&overview=closed&s=4#tabs

Contissa, Giuseppe, and Francesca Lagioia, Giovanni Sartor. “The Ethical Knob: ethically-customisable automated vehicles and the law.” Artificial Intelligence and Law, 25.3, 2017, pp. 365-378.

Davnall, Rebecca. “Solving the Single-Vehicle Self-Driving Car Trolley Problem Using Risk Theory and Vehicle Dynamics.” Science and Engineering Ethics, vol. 26, no. 1, Springer Netherlands, 2019, pp. 431–49.

Faulhaber, Anja K., and Anke Dittmer, Felix Blind, Maximilian A. Wächter, Silja Timm, Leon R. Sütfeld, Achim Stephan, Gordon Pipa, Peter König. “Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles.” Science and engineering ethics, 25.2, 2018, pp. 399-418.

The Federal Ministry of Transport and Digital Infrastructure. “Ethics Commission Automated and Connected Driving Report 2017,” 2017, https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Foot, Philippa. “The problem of abortion and the doctrine of the double effect.” Oxford review, 1967.

Fournier, Tom. “Will my next car be a libertarian or a utilitarian?: Who will decide?” IEEE Technology and Society Magazine, 35.2, 2016, pp. 40-45.

Frank, Darius-Aurel, and Polymeros Chrysochou, Panagiotis Mitkidis, Dan Ariely. “Human decision-making biases in the moral dilemmas of autonomous vehicles.” Scientific Reports, vol. 9, no. 1, NATURE PUBLISHING GROUP, 2019, pp. 13080–13119.

Freitas, Julian De, and Sam E. Anthony, Andrea Censi, George A. Alvarez. “Doubting Driverless Dilemmas.” Perspectives on Psychological Science, vol. 15, no. 5, SAGE Publications, 2020, pp.1284–1288.

Frison, Anna-Katharina, and Philipp Wintersberger, Andreas Riener. “First person trolley problem: Evaluation of drivers’ ethical decisions in a driving simulator.” Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications, 2016, pp. 117-122.

Furukawa, Osamu. “Development of Automated Driving Technology: Its History and Direction for Practical Application” Grand Prix Publishing Co. 2019.

Gill, Tripat. “Blame it on the self-driving car: how autonomous vehicles can alter consumer morality.” Journal of Consumer Research, 47.2, 2020, pp. 272-291.

Gogoll, Jan, and MÜLLER, Julian F Müller. “Autonomous cars: in favor of a mandatory ethics setting.” Science and engineering ethics, 23.3, 2017, pp. 681-700.

Goodall, Noah J. “Away from trolley problems and toward risk management,” Applied Artificial Intelligence, 30.8, 2016, pp. 810-821.

Gurney, Jeffrey K. “Crashing into the Unknown: An Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law.” Albany Law Review, vol. 79, no. 1, Albany Law School, 2015, p. 183.

Hern, Alex. “Self-driving cars don’t care about your moral dilemmas.” The Guardian, 22 Aug, 2016, https://www.theguardian.com/technology/2016/aug/22/self-driving-cars-moral-dilemmas

Himmelreich, Johannes. “Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations.” Ethical Theory and Moral Practice, vol. 21, no. 3, Springer, 2018, pp. 669–684.

HONDA. “Launch of the new LEGEND equipped with Honda Sensing Elite” 4 Mar. 2021, https://www.honda.co.jp/news/2021/4210304-legend.html.

Hosaka, Akio, and Aoki keiji and Sadayuki tsugawa. “Automated driving (2nd edition) – System structure and elemental technologies –.” Morikita Publishing Co. 2020.

The Institute of Electrical and Electronics Engineers website. “IEEE Code of Ethics” Jun.2016, https://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/corporate/ieee-code-of-ethics.pdf

Kallioinen, Noa, and Maria Pershina, Jannik Zeiser, Farbod Nosrat Nezami, Gordon Pipa, Achim Stephan, Peter König. “Moral Judgements on the Actions of Self-Driving Cars and

Human Drivers in Dilemma Situations From Different Perspectives." Frontiers in Psychology, vol. 10, Frontiers Research Foundation, 2019, pp. 2415–2415.

Kawashima, Shigeo, and Satoshi Kitamura, Yasufumi Shibanai. "Attitudes toward the 'trolley problem' and other issues of automated vehicles: through a questionnaire survey of people living in Japan." Presentation Material for the 2017 Annual Conference of the Japan Society for Socio-Information Studies, 2017.

Krügel, Sebastian, and Matthias Uhl. "Autonomous Vehicles and Moral Judgments Under Risk." Transportation Research. Part A, Policy and Practice, vol. 155, Elsevier Ltd, 2022, pp. 1–10.

Lin, Patrick. "Why ethics matters for autonomous cars. In: Autonomous driving." Springer, Berlin, Heidelberg, 2016, pp. 69-85.

Lundgren, Björn. "Safety requirements vs. crashing ethically: What matters most for policies on autonomous vehicles." AI & Society, vol. 36, no. 2, Springer London, 2020, pp. 405–415.

Maxmen, Amy. "A moral map for AI cars." Nature, 562.7728, 2018, pp. 469–470.

McFarland, Matt. "Uber self-driving car kills pedestrian in first fatal autonomous crash" CNN, 19 Mar. 2018, https://money.cnn.com/2018/03/19/technology/uber-autonomous-car-fatal-crash/index.html?iid=EL

Mirnig, Alexander, and Alexander Meschtscherjakov. "Trolled by the trolley problem." proceedings of the 2019 CHI conference on human factors in computing systems, ACM, 2019, pp. 1–10.

Morita, Tamayuki, and Shunsuke Managi. "Demand and Social Dilemmas Created by Automated Driving Cars," Research Institute of Economy, Trade and Industry, 2018.

The National Highway Traffic Safety Administration. "Automated Driving System 2.0: A Vision for Safety." 12 Sept. 2017, https://www.nhtsa.gov/document/automated-driving-systems-20-voluntary-guidance

The National Highway Traffic Safety Administration. "Federal Automated Vehicles Policy." 20 Sept. 2016, https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016

The National Highway Traffic Safety Administration website. "Overview of Motor Vehicle Crashes in 2020." Mar. 2022, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266

The National Highway Traffic Safety Administration. "Preliminary Statement of Policy Concerning Automated Vehicles." May. 2013, https://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

The National Police Agency of Japan website. “The Number of Traffic Accident Fatalities in 2021.” 4 Jan. 2022, https://www.npa.go.jp/news/release/2022/20220104001jiko.html

The National Science & Technology Council and the United State Department of Transportation. “Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0.” 8 Jan. 2020, https://www.transportation.gov/av/4.

Nikkei. “Honda to launch the world’s first “Level 3” Automated driving system on May 5.” 4 Mar. 2021, https://www.nikkei.com/article/DGXZQODZ032HJ0T00C21A3000000/

Nikkei. “The Japanese government has approved a draft law to lift the ban on ‘Level 4’ automated driving.” 4 Mar. 2022, https://www.nikkei.com/article/DGXZQOUE034G70T00C22A3000000/

Novak, Thomas P. “A Generalized Framework for Moral Dilemmas Involving Autonomous Vehicles: A Commentary on Gill.” The Journal of Consumer Research, 47(2), 2020, pp. 292–300.

Nyholm, Sven R., and Jillles Smids. “The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem?” Ethical Theory and Moral Practice 19 (5), 2016, pp. 1275–89.

Olson, Edwin. “Trolley Folly.” May Mobility, 14 Nov, 2018, https://medium.com/may-mobility/trolley-folly-fcbd181b7152

The Open Roboethics Initiative. “If death by autonomous car is unavoidable, who should die? Reader poll results.” 23 Jun. 2014, https://robohub.org/if-a-death-by-an-autonomous-car-is-unavoidable-who-should-die-results-from-our-reader-poll/

Page, Misconceptions. “Top misconceptions of autonomous cars and self-driving vehicles.” 2016.

The Prime Minister’s Official Residence Website. “Public-private ITS initiative and roadmap: Past efforts and basic concept of future ITS initiative.” Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society and Public-Private Data Utilization Promotion Strategy Council, 15 Jun. 2021, https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20210615/roadmap.pdf

Radun, Igor, and Jenni Radun, Jyrki Kaistinen, Jake Olivier, Göran Kecklund, Tòres P. G. Theorell. “Endangering yourself to save another: A real life ethical dilemma.” Transportation Research. Part F, Traffic Psychology and Behaviour, 64, 2019, pp. 318–322.

Russell, Stuart, and Daniel Dewey, Max Tegmark. “Research priorities for robust and beneficial artificial intelligence.” Ai Magazine, 36.4, 2015, pp. 105-114.

The SAE International J3016. “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicle” Sep. 2016, https://www.sae.org/standards/content/j3016_201609/

Sato, Hideaki. “Automated vehicles and the trolley problem.” Chuo Gakuin University Human and Nature Series, 48, 2020, pp. 21-54.

Sekine, Michiaki, and Kaneo Hiramatsu. “Definition of automated driving technology and status of study on items for international standardization at the United Nations Autonomous Driving Subcommittee (ITS/AD)” Summary of the 2017 lecture at the Traffic Safety and Environment Research Institute Forum, 2017, pp. 47-50, https://www.ntsel.go.jp/Portals/0/resources/forum/2017files/1201_1430.pdf

Smith, Brent. “Personality facets and ethics positions as directives for self-driving vehicles.” Technology in Society, 57, 2019, pp. 115–124.

Smith, Bryant W. “The Trolley and the Pinto.” Texas A&M Law Review 4 (2), 2017, pp. 197–208.

The Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society website. “Outline of System Development for Automated Driving” 17 Apr. 2019, http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20180413/auto_drive.pdf

Sütfeld, Leon R., and Richard Gast, Peter König, Gordon Pipa. “Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure.” Frontiers in behavioral neuroscience 11:122, 2017.

Taylor, Michael. “Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians.” Car and Driver, 7 Oct, 2016. https://www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/

Thomson, Judith Jarvis. “Killing, letting die, and the trolley problem.” The monist, 59.2, 1976, pp. 204-217.

Thomson, Judith Jarvis. “The trolley problem.” 1985.

Thomson, Judith Jarvis. “Turning the trolley.” Philosophy & Public Affairs, 36.4, 2008, pp. 359-374.

The United State Department of Transportation. “Preparing for The Future of Transportation: Automated Vehicles 3.0” 4 Oct. 2018, https://www.transportation.gov/av/3

Wiseman, Yair, and Ilan Grinberg. “The Trolley Problem Version of Autonomous Vehicles.” The Open Transportation Journal 12 (1), 2018, pp. 105–13.

The World Health Organization. “Road traffic injuries” 21 Jun. 2021, https://www.nhtsa.gov/press-releases/2020-fatality-data-show-increased-traffic-fatalities-during-pandemic

Wu, Stephen S. “Autonomous Vehicles, Trolley Problems, and the Law.” Ethics and Information Technology, vol. 22, no. 1, Springer Netherlands, 2019, pp. 1–13.

Zhu, Anrun, and Shuangqing Yang, Yunjiao Chen, Cai Xing. “A moral decision-making study of autonomous vehicles: Expertise predicts a preference for algorithms in dilemmas.” Personality and Individual Differences 186, 2022, p. 111356.