

# Automotive Ethics 2.0

Professor Wolf Schäfer  
Department of Technology &  
Society

DTS Lecture Series

December 11, 2019

Automated Driving Systems (ADS) will operate on land, in the oceans as well as in air and space. Their first major appearance will be self-driving cars and trucks on highways and in cities. Building supersafe robotic vehicles is a great engineering promise, however, its most underdeveloped element is the design of automotive ethics. Enabling these vehicles for critical decision-making in edge cases is crucial, but also the most obscure part of their moral machinery. ADS programming must prepare for accidents, systemic failures, hacking attacks, and dilemma situations, yet engineering lacks a universal machine ethics and an overarching moral code for AI-driven systems. Various ethical theories – libertarianism, utilitarianism, Kantianism, for example – are available, yet an ADS implementation of any of them would appear arbitrary. The design of morally valid, universal ADS control systems is further complicated by different moral preferences in large Western, Eastern, and Southern cultural clusters. Hence, responsible engineering must begin to address automotive ethics now, before the massive advent of ADS.

*responsible  
engineering  
education goes  
**beyond a 1-  
credit ethics  
course***

1. **How Much Is Too Much Ethics?**
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov Toward Morality By Math
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

# How Much Ethics?

**Nov. 6, 2019, email to a DTS staff member from a CEAS GPD** (Graduate Program Director; full professor in Mechanical Engineering):

Your department requested an approval for the same EST 502 Ethics course in last spring.

At that time, I raised a concern that the subject being 3 credit course. I supposed it was not approved then (?)

Anyway, I do have a reservation for offering “Ethics” as a full 3-credit course although it is a very important matter and everyone should understand it.

**Ethics is not an academic subject in engineering/science/technology.** As such the course should be offered as a “training” course as we need to take online ethics video (JCOPE) or at most 1 credit course.

## **“Ethics” not supposed to exceed 1 credit ?**

**Nov. 7, 2019, 4:32 PM, DTS GPD response:**

The ethics course is integral to our MS program in CEAS. We teach management of technology and engineering (some to graduates of your program), and lord knows we want our managers to have a strong understanding of ethics and to make decisions based on ethical principles. We think the material we cover in the course and the work we ask of the students warrants three credits. I hope you will re-consider your comment in this context.

Thanks for your thoughtful response -- even as I hope you will change it.

**Nov. 11, 2019, 10:15 AM, CEAS GPD response** (full professor in Applied Mathematics and Statistics):

I approve EST 519, but disapprove EST 502 for the same reason ...

In AMS, for example, AMS 500, a zero-credit ethics course, introduces students to the major issues in the ethics of science and research.

This course meets for 1 hour for 8 weeks, and each professor is assigned to teach it for one week.

I agree ... that an ethics course is not supposed to exceed 1 credit.

## ***“Ethics” not an academic subject in engineering/science/technology ?***

**Nov. 11, 2019, 1:33 PM, DTS GPD response:**

I am glad you and your department are engaged in ethical considerations. I'm not sure that 8 hours are sufficient to cover everything that's needed, however, even for very efficient mathematical thinkers. Financial modelers responsible for the 2008 crash might have wanted to consider the ethics of mortgage bundling and reliance on models such as VaR; mathematics and math-related industries face issues with under-representation by particular sets of people; and growing reliance on algorithms while seemingly providing a level of objectivity as to how decisions are made also can include a good ration of bias in how those algorithms are constructed. These are deep and pressing issues for modern society and are not easily resolved. You may disagree with how I represent the three exemplar issues, but that we may not agree provides support for more robust investigations of the topic, I think. Does that mean your course needs to be expanded -- well, perhaps not, but a deeper dive into the subject [of ethics] in CEAS would not be a bad thing, in my opinion. And my department deeply believes this.

**Nov. 11, 2019, 5:24 PM, CEAS GPD response:**

Hi, Based on your explanation, I approve EST 502. It seems like your program has a lot more topics to be covered in the ethics course than ours.

# *automotive ethics from 1.0 to 2.0*

1. How Much Is Too Much Ethics?
2. **What Is Automotive Ethics?**
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov Toward Morality By Math
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

## Searching for Automotive Ethics

In April 2019, I googled “automotive ethics” and received over 62 million answers in less than 1 second. The top answer was: **“to perform high-quality repair service at a fair and just price”** – the first commandment of the Code of Ethics of the American Automotive Service Association (ASA).

Yesterday, I checked again. The top results still pointed to the ASA Code of Ethics for mechanics and automobile dealers – what we may call **Automotive Ethics 1.0**.

**“Automotive Ethics – Stony Brook University”** came also up on the first results page both yesterday and in April. This was/is pretty good since we had listed our research initiative only in March 2019.

These search results show

- the automotive experience of the last century and its associated concerns are still very much with us.
- our understanding of Automotive Ethics as **Automotive Ethics 2.0** needs a lot more research and much clarification.

Daniene Byrne\* and I are working on that.

## Code of Ethics

The owners and managers of automotive service businesses that belong to the Automotive Service Association (ASA) agree to adhere to a Code of Ethics. ASA's Code of Ethics is the automotive service industry's standard for professional business practices.

- To perform high quality repair service at a fair and just price.
- To use only proven merchandise of high quality distributed by reputable firms.
- To employ the best skilled technicians obtainable.
- To furnish an itemized invoice for fairly priced parts and services that clearly identifies any used or remanufactured parts. Replaced parts may be inspected upon request.
- To have a sense of personal obligation to each customer.
- To promote good will between the motorist and members of the association.
- To recommend corrective and maintenance services, explaining to the customer which of these are required to correct existing problems and which are for preventive maintenance.
- To offer the customer a price estimate for work to be performed.
- To furnish or post copies of any warranties covering parts or services.
- To obtain prior authorization for all work done, in writing, or by other means satisfactory to the customer.
- To notify the customer if appointments or completion promises cannot be kept.
- To maintain customer service records for one year or more.
- To exercise reasonable care for the customer's property while in our possession.
- To maintain a system for fair settlement of customer's complaints.
- To cooperate with established consumer complaint mediation activities.
- To uphold the high standards of our profession and always seek to correct any and all abuses within the automotive industry.
- To uphold the integrity of all members of the Automotive Service Association.

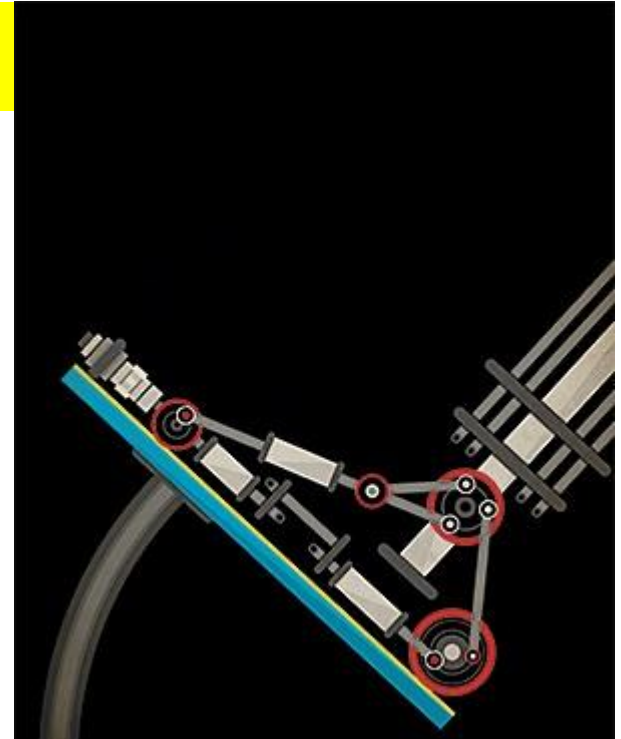
# Automotive Ethics 2.0 = AI + E

**DTS at Stony Brook adds the E of ethics to the artificial intelligence of AVs and AEVs: AI + E.**

Electric vehicles (EVs) are already on our streets. Automated\* Vehicles (AVs) and Automated Electric Vehicles (AEVs) are coming. The latter are no longer human-driven but steered by Artificial Intelligence (AI).

The difference between Automotive Ethics 1.0 and 2.0 is between an **aspirational Code of Ethics** like the Ten Commandments, which human actors *should* follow, and a **preprogrammed algorithmic ethics** that Automated Driving Systems (ADS) *will* follow.

\* **A footnote on our terminology:** We distinguish between “automated” and “autonomous.” A Nissan engineer nailed the reason why when he said: “A truly autonomous car would be one where you request it to take you to work and it decides to go to the beach instead.”





*engineering has  
become much  
too important  
to be left to the  
engineers*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
- 3. The DTS Motto and What It Means/Not Means**
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov Toward Morality By Math
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

## Mushroom Clouds at the Origin of Our Motto

Natural science lost its purity in the First and Second World Wars: Chemistry in WWI by enabling chemical warfare; physics in WWII by building the first nuclear bombs.

The head administrator of the Manhattan Project – James Bryant Conant – concluded after Hiroshima and Nagasaki: **“Science is much too important to be left to the scientists.”**

Today, engineering has eclipsed the attraction of physics, but its brand as a purely problem-solving endeavor has also darkened. Now, engineering solutions are judged by their achievements *and* their consequences. Biased machine-learning, deep fake videos, privacy loss, and Cambridge Analytica distortions have tarnished engineering.

Thus, ***engineering has become much too important to be left to the engineers.*** Engineering must go *far beyond* merely technical solutions. Open up, leave its comfort zone, and assess the social and natural impacts of its innovations.



## What Does Opening Up Mean/Not Mean?

**DTS has a binocular focus; we look at technology AND society.**

Algorithms that select or deselect mates, job-seekers, or drone targets – DTS faculty and students are likely to study their technical achievements *and* non-technical impacts.

- Opening up means *all* engineering students should take a 3-credit ethics course.
- Automotive ethics means that we consider the employment consequences of self-driving trucks.
- Yet opening up **does not mean technical expertise should be a secondary qualification.**

Take Boeing's recent 737 MAX crashes. In 1996, Boeing acquired McDonnell Douglas for 13 billion dollars. After the merger, "Boeing went from being led by engineers to being led by business executives driven by stock performance."\*

To prevent this deadly mistake, **DTS marries technical expertise with societal impact studies and values both equally.**



*disruptive  
change in ever  
shorter time  
spans*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
- 4. The Perils of Growth and the Promise of the 4th Industrial Revolution**
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov Toward Morality By Math
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

## Humanity's Growth

### Four Global Energy Transitions:

1. The domestication of fire
2. The domestication of plants and animals
3. The domestication of fossil fuels
4. The domestication of nuclear power

These four revolutionary transitions have increased Earth's carrying capacity and triggered subsequent population explosions.

- **1 million** hunter-gatherers 10,000 BCE
- **1 billion** farmers 1800 CE
- **2.5 billion** industrialists in the semi-industrialized world of 1950
- **7.7 billion world population now**
- **10 billion** estimated to live in a fully industrialized world by 2050
- **100 billion** with nuclear fusion in 3000? Not likely, but possible.

## and Shrinking Adaptation Time

The rub of humanity's inventive history:  
**Disruptive change has to be weathered in ever shorter time spans.**

- Hunter-gatherers had over **1 million years** or 90% of human history
- Agriculturalists had **12,000 years**
- Industrialists have had **200 years** so far
- The nuclear age dawned in the **1940s** in my life time

Humanity's adaptation window has shrunk by six orders of magnitude, from  $10^6$  to less than  $10^1$  years. **Are we prepared for technology-driven change in ever shorter time spans?**

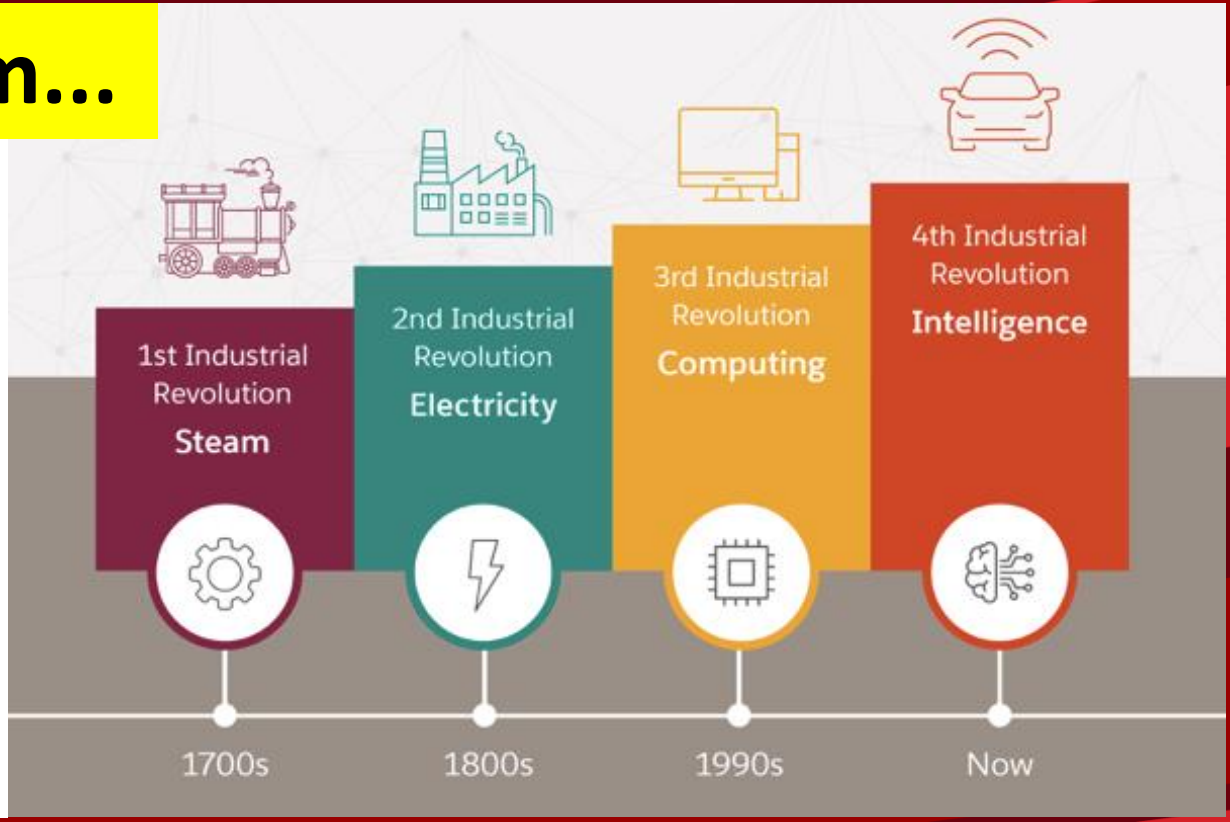
# No...Far from...

**However, there is hope.**

The Industrial Revolution, which was a local affair when it started, is going global now and becoming intelligent.

**The Fourth Industrial Revolution** (Industry 4.0) offers powerful help. Combining human and machine intelligence, it promises to rapidly tackle complex problems.

Intelligent cyber-physical systems, such as robo-cars and -trucks, can herald supersafe travel and efficient transportation.



*a global  
technoscientific  
civilization and  
premature  
claims...*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
- 5. Big Assumptions and Baby Steps**
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov To Morality By Math
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

# Big Assumptions

1. Humanity is building a global technoscientific civilization, which I have named **Pangaea Two**.\*
2. Cyber-physical systems, which are the signature-feature of Industry 4.0, support Pangaea Two.
3. Smart design is the core business of Industry 4.0. Everything, from genetically engineered fish to cars without steering wheels, is design-based.
4. Breakthrough technologies, such as AI, biotechnology (CRISPR), nanotechnology, fifth-generation wireless (5G), and quantum computing, permit design across physical, biological, national, and disciplinary borders.
5. Designing cyber-physical systems across political, natural, and academic boundaries empowers inter- and transdisciplinary team science.
6. SBU, CEAS, and DTS are harnessing team science in Vertically Integrated Projects (VIP).
7. Different political and regulatory cultures will impact the design and release of AI controlled vehicles.
8. Technology entrepreneurs in the US are chasing the next disruptive innovation.
9. Most disruption-promising projects proceed without paying close attention to ethical responsibility, social harmony, and global sustainability.
10. AVs and AEVs warrant the inclusion of a holistic moral machine, even in a predominantly deregulatory environment like the US.
11. AI-driven vehicles will be held accountable for their driving decisions, no matter how much they will lower the global burden of road traffic injuries and death.
12. AI + E enables **controlled disruption**.



# Controlled Automotive Disruption

The traditional automotive industry is inclined towards *controlled* disruption. As one of the world's leading economic sectors by revenue, it competes on a global scale and will catch up with pioneering disruptors like Tesla.

For self-preservation and continued leadership, the established automotive industry will pay careful attention to

- national and international road safety regulations,
- social and ecological sustainability, as well as
- the demands of AI + E.

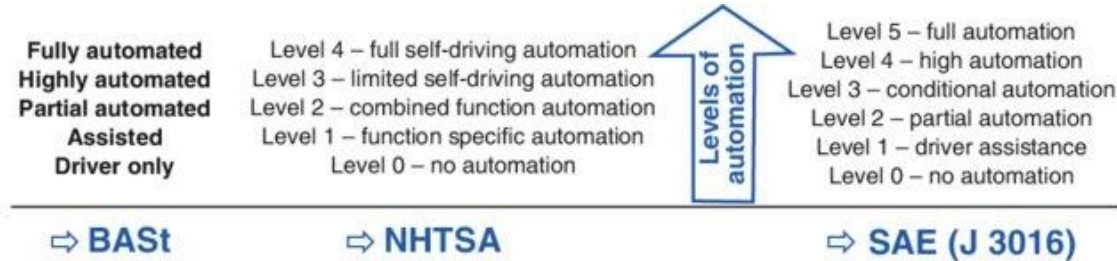
**My Forecast:** The large-scale introduction of AEVs will disrupt established patterns of transportation and mobility worldwide, yet also save countless lives. Highly effective, automatic safety measures will significantly reduce motor vehicle collisions<sup>1</sup> and curb the pollution caused by fossil fuel vehicles.<sup>2</sup> Additional benefits can be expected in the area of raw material procurement.<sup>3</sup>

<sup>1</sup> According to the WHO (*Global Status Report on road safety 2018*), road traffic injury is the **8th leading cause of death** for all age groups and still increasing. The current global burden of road traffic deaths is **1.35 million people**.

<sup>2</sup> The UK and France have set the end of gas and diesel vehicles by 2040; Norway has decreed its respective deadline for 2025.

<sup>3</sup> Presently, electromobility based on lithium-ion batteries incurs brutal socio-natural costs from unregulated mining of raw materials (child labor, steep environmental degradation, huge health and safety hazards).

## Baby Steps from Zero to Full Automation



Levels of automation defined by the German *Bundesanstalt für Straßenwesen* (BASt), the US National Highway Traffic Safety Administration (NHTSA), and the international Society of Automotive Engineering (SAE).

One can say, *The race to increase the levels of vehicle automation is on*. One can also say, *The AEV revolution is underway*. But one cannot say the paradigm shift to AI-driven vehicles is all but accomplished. However, that is what Elon Musk said in April 2019.<sup>1</sup>

*Consumer Reports* and the industry-coalition *Partners for Automated Vehicle Education* (PAVE) immediately declared such claims premature and “not backed up by the data.”<sup>2</sup>

Currently, commercially available automated cars are only partially automated; none exceeds SAE level 2. Full automation is still far away.

<sup>1</sup> See Russ Mitchell. “Elon Musk Claims a Million Teslas Will Drive Themselves in a Year. Safety Advocates Have Concerns.” [latimes.com](https://www.latimes.com), 22 April 2019.

<sup>2</sup> “Tesla Must Prove Safety Before Claiming ‘Self-Driving’ Ability.” [Consumer Reports](https://www.consumerreports.com), April 22, 2019. See also Bill Visnic. “PAVE Coalition Formed,” January 8, 2019, [sae.org/news/2019/01/pave-coalition-announcement](https://www.sae.org/news/2019/01/pave-coalition-announcement).

## *a misguided moral machine experiment*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
- 6. Automotive Ethics 2.0 Now! But Not the Wrong Way**
7. From Asimov Toward Morality By Math
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

## Zoom Ahead With Ethics

The best time to accelerate research into Automotive Ethics 2.0 is **Now**, ahead of the widespread use of levels 3 to 5 vehicles.

The awareness of ethical problems in engineering and applied sciences tends to rise when applications become ubiquitous. Take facial recognition technology. What was an innovative idea and tempting AI challenge turned into a Black Mirror-type issue after it was adopted wholesale by the police in China and elsewhere.

We can do better. In order to move from a *reactive* to a *proactive* stance, we must investigate the moral hazards of emerging AI technologies **before** they massively arrive.

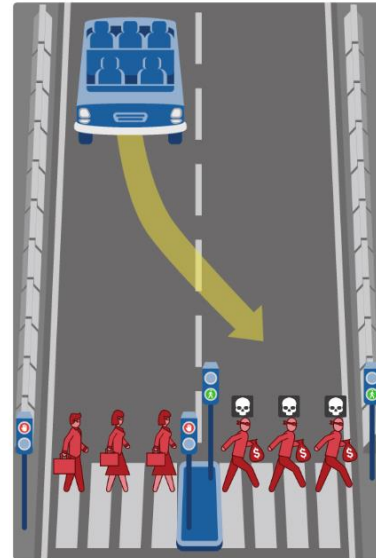
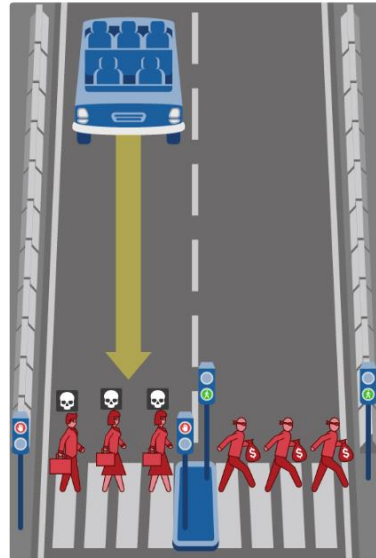
If we wait until potential risks have become realized, our options and associated time frames will be severely restricted. Hence, the time to accelerate research into automotive ethics is **before** ADS are in widespread use.

# How Not To Do Automotive Ethics

The brakes of an automated vehicle are out and it has 2 options. *Left:* Kill 1 male and 2 female executives. *Right:* Kill 3 criminals.

Which group should the car's AI kill?

*Nota bene:* The 3 executives are disregarding a red light, whereas the 3 criminals cross properly at green.



The graphic presents one of the many scenarios of MIT's "Moral Machine Experiment."

People from over 233 countries and territories provided over 40 million decisions in 10 languages.<sup>1</sup>

The MIT experiment is a good example for the harnessing of mass collaboration for an online research project.

[moralmachine.mit.edu](http://moralmachine.mit.edu)

# A Bias-Based Approach

**The Moral Machine of an AV should not take social properties into account.**

Social facts – rich or poor, slim or overweight, occupation, religion, gender and sexual orientation – should be off limits in the determination of lethal action.

Yet by doing just that – invoking social properties – the MIT experiment triggered the targeting of supposedly “lesser” people. By testing people’s biases, it laid open global structures of discrimination and profiling. That was good social science, but invited bad ethics as well.

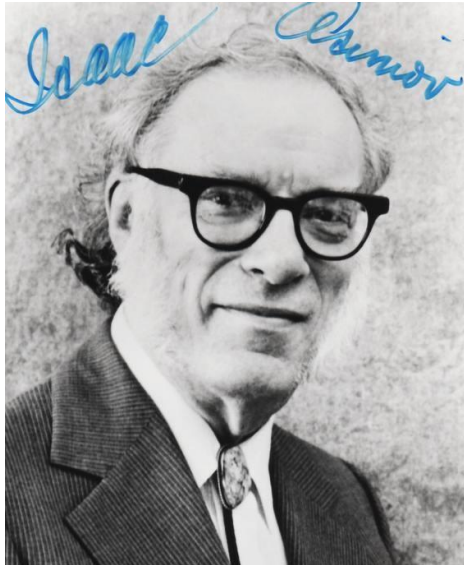
*One more thing:* If the AV knows that the three men who are crossing the street legally at green are “criminals” (have criminal convictions), the vehicle’s AI must have real time access to police records and other relevant databases.

As far as I am concerned, **responsible AI + E should not be allowed to make lethal decisions based on police or other surveillance data.**

*from morality  
by intuition to  
morality by  
math via  
science fiction*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
- 7. From Asimov Toward Morality By Math**
8. Solving AI + E Math Problems
9. The Conundrum of Lethal Moral Math

# Asimov's Three Laws



**Normative distinctions**, such as good or bad, right or wrong, can be *determined* by democratic consent or authoritarian diktat, whereas **technical values**, such as correct or incorrect, must be *discovered* by science and engineering. One of the first attempts to resolve this notorious tension between values and facts for intelligent robots was formulated in 1942 by science fiction writer **Isaac Asimov** in his **Three Laws of Robotics**:

1. *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
2. *A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.*
3. *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*<sup>1</sup>

<sup>1</sup> See Andrew Liptak. "Isaac Asimov and the Three Laws of Robotics." [Kirkus Reviews](#), March 14, 2013.

Asimov extends humanity's religious and philosophical rules-tradition – which includes **Do No Harm** – from fellow human beings to robots. Yet what has worked for humans (sort of), does not work for robots. To expose this flaw, philosopher Derek Leben has asked, What exactly counts as "harm"? Does lying, trespassing, or intrusion of a person's privacy constitute "harm"? What if "harm" is only *likely*? Can't any action or inaction result in some kind of "harm"? And what happens when all available choices are harmful? Should the robot stop dead? To avoid these pitfalls, **functional AI + E has to specify programmable thresholds to save robots from debilitating paralysis.**<sup>2</sup>



**Recent work on automotive ethics** has applied Asimov's laws to AVs:

1. *An automated vehicle should not collide with a pedestrian or cyclist.*
2. *An automated vehicle should not collide with another vehicle, except where avoiding such a collision would conflict with the First Law.*
3. *An automated vehicle should not collide with any other object in the environment, except where avoiding such a collision would conflict with the First or Second Law.*<sup>1</sup>

The term of art philosophers use for strict obligations like those above is **deontic**. (Kantian ethics would be an example of a deontic moral theory.) The problem with deontology is its moral rigor. **Deontological rules leave no wiggle room for negotiation or situational adaptation.**

Consequentialism stands opposite deontology and argues that **any action or inaction that produces a good outcome is morally right**. (American pragmatist philosopher John Dewey could be called a consequentialist).

Gerdes and Thornton use both frameworks – deontology plus consequentialism – to impart optimal ethical behavior into automotive AI. To achieve this goal, they map both theories onto mathematical programs that control the vehicle's decision-making.

*collision  
programming:  
hairy snags  
and killer  
features*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov Toward Morality By Math
- 8. Solving AI + E Math Problems**
9. The Conundrum of Lethal Moral Math

# Unintended Consequences of Moral Math

When you program an automotive robot for *responsible* driving, **expect the unexpected!** Here's an example:

A fully automated SAE Level 5 vehicle cannot avoid crashing into one of two cyclists, yet it can select which one: the one, who is wearing a helmet, or the other one, who is not. What is the “responsible” choice? Hitting the protected or the unprotected cyclist?

The answer seems to be: *Target the person that is more likely to survive.* Hence, the responsibly programmed car hits the guy with the helmet and not the other who might die. Now, the AV has saved a life, yet also penalized lawful and prudent road behavior.

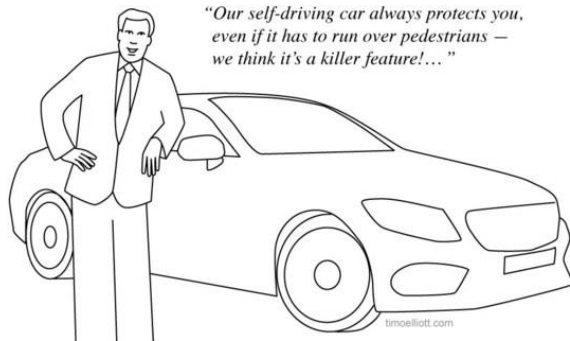
Furthermore, this crash-optimization program has now set the stage for more cyclists to forego helmets, because doing so has become safer with respect to the optimal targeting choice of the automated car.<sup>1</sup>



**Writing moral values into the algorithms of automated cars is necessary and unavoidable, but also fraught with uncertainty and unintended consequences.**

# A Sociopolitical Roadblock for AI + E

**Optimal control theory** is a mathematical method “directly analogous to consequentialist approaches in philosophy” (Gerdes & Thornton). It translates the ethical implications of all actions and inactions into **cost functions**, which in turn allow the software controller to reward desired and penalize undesired actions. This method can guide a rocket to its destination and give automated vehicles smooth steering and safe braking. Calculating the cost of property damage versus personal injury, or the difference between occupant versus pedestrian protection, is in its reach. **The problem is neither mathematical nor philosophical, but rather the development of acceptable cost functions.**



In 2016, a Daimler manager was quoted saying, “If you know you can save at least one person, at least save that one. Save the one in the car ... If all you know for sure is that one death can be prevented, then that’s your first priority.” The media exploded. *Car and Driver* headlined, “Self-Driving Mercedes-Benzenes Will Prioritize Occupant Safety over Pedestrians.”\* There’s a lesson here: “Morality by math” (Patrick Lin) is not natural; it has to be researched, taught, and explained. **The mathematical morality of control algorithms must be made transparent and understandable.**

# Watch How Math Morality Works

See at [youtube.com/watch?v=NG\\_O4RyQqGE](https://youtube.com/watch?v=NG_O4RyQqGE)

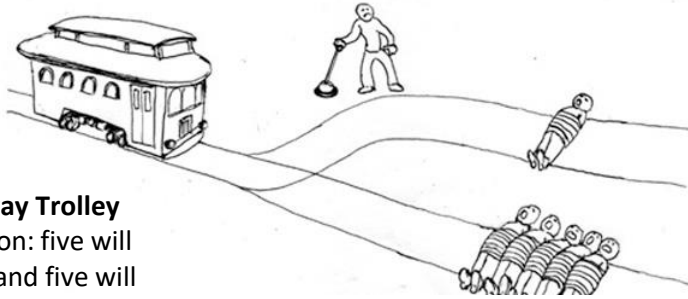
The video shows *how* a **switch from consequentialist to deontic logic** prevents the AVs collision with two cyclists. Instead of making the potential collision a million times more costly than a lane change, the programmer has put a firm constraint on the optimization process illustrated by the purple crossbar blocking both lanes. This bar is an encoded hard constraint, that is, a mathematical command instructing the AI: ***Do Not Pass These Cyclists.***



# *the brave illusion of total safety*

1. How Much Is Too Much Ethics?
2. What Is Automotive Ethics?
3. The DTS Motto and What It Means/Not Means
4. The Perils of Growth and the Promise of the 4th Industrial Revolution
5. Big Assumptions and Baby Steps
6. Automotive Ethics 2.0 Now! But Not the Wrong Way
7. From Asimov Toward Morality By Math
8. Solving AI + E Math Problems
- 9. The Conundrum of Lethal Moral Math**

## Lethal Moral Math



This is the famous/infamous<sup>1</sup> **Runaway Trolley Problem** depicting a dilemma situation: five will live and one will die, or one will live and five will die. If you take the human at the switch away and automate the trolley, the vehicle's AI is asked to make the lethal decision. Guess what!

*Nobody will die, our programmer may say, because my car gets the deontic crossbar and stops before the split in the track.* – Well, that answer avoids both the problem and the dilemma. The point of the ethical dilemma is the question: Which alternative *shall happen* (selected by the AI) when a lethal accident *will happen* (caused by the AV)?

Original Equipment Manufacturers (OEM) and AI engineers are reluctant to engage with thought experiments that involve lethal dilemmas. They are chasing the science fiction goal of total safety in which accidents have become a thing of the past. But dilemma-style situations will happen, robots will be hacked, fail-safe systems will fail and overrides freeze. Hence:

**AI + E cannot avoid to investigate and discuss possible lethal moral math problems.**

# CHANGING THE GREAT BIG WORLD

The slogan of the SBU PowerPoint template (left) seems to celebrate change for the sake of change.

**Automotive Ethics 2.0 seeks Supersafe Automobility for All and Advanced Ethics Education for Engineers.**