

Data and Society

Data and Ethics – Lecture 19

4/11/20

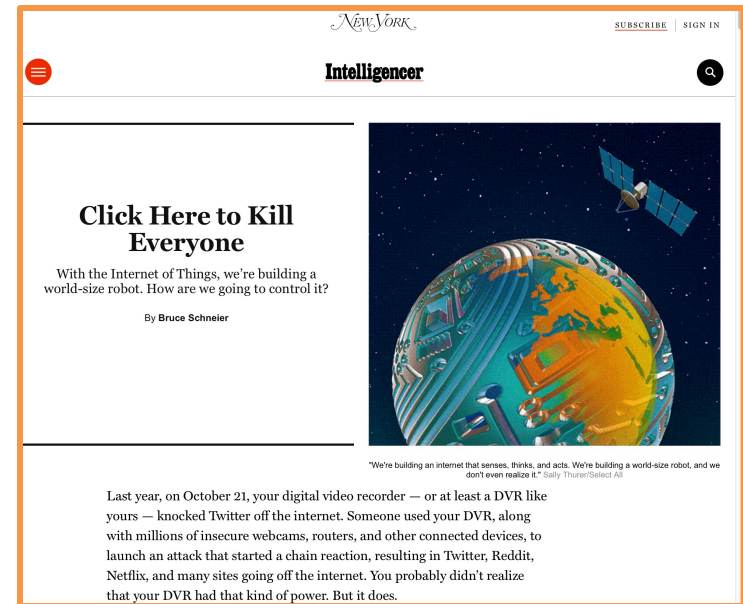
Today's Class

- Choice assignment due 11:59 p.m. April 15 / Instructions in Lecture 17
- Guest speaker **April 15:**
Bruce Schneier, Harvard U.
- Lecture / Discussion
- Student Presentations



Reading for April 15

- **“Click Here to Kill Everyone”**, Bruce Schneier, New York Magazine, <https://nymag.com/intelligencer/2017/01/the-internet-of-things-dangerous-future-bruce-schneier.html>



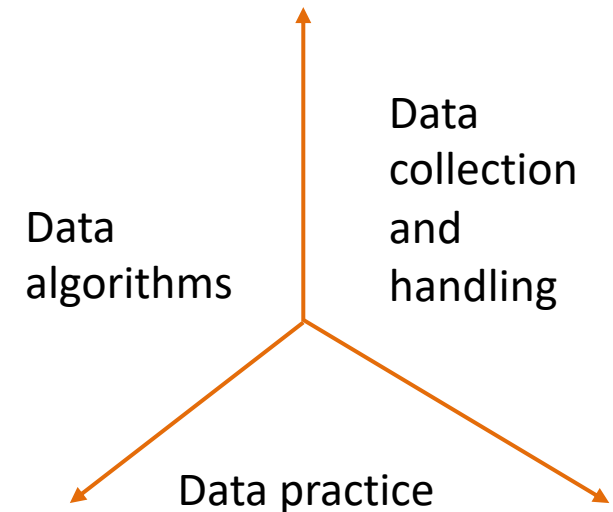
Date	Topic	Speaker	Date	Topic	Speaker
1-25	Introduction	Fran	1-28	The Data-driven World	Fran
2-1	Data and COVID-19	Fran	2-4	Data and Privacy -- Intro	Fran
2-8	Data and Privacy – Differential Privacy	Fran	2-11	Data and Privacy – Anonymity / Briefing Instructions	Fran
2-15	NO CLASS / PRESIDENT’S DAY		2-18	NO CLASS	
2-22	Legal Protections	Ben Wizner	2-25	Data and Discrimination 1	Fran
3-1	Data and Discrimination 2	Fran	3-4	Data and Elections 1	Fran
3-8	Data and Elections 2	Fran	3-11	NO CLASS / WRITING DAY	
3-15	Data and Astronomy (Op-Ed due)	Alyssa Goodman	3-18	Data Science	Fran
3-22	Digital Humanities	Brett Bobley	3-25	Data Stewardship and Preservation	Fran
3-29	Data and the IoT	Fran	4-1	Data and Smart Farms	Rich Wolski
4-5	Data and Self-Driving Cars	Fran	4-8	NO CLASS – WELLNESS DAY	Fran
4-12	Data and Ethics	Fran	4-15	Cybersecurity	Bruce Schneier
4-19	Data and Dating	Fran	4-22	Digital Rights in the EU and China	Fran
4-26	Tech in the News	Fran	4-29	NO CLASS	Fran
5-3	Wrap-up / Discussion				

Lecture

- Data and Ethics
- The Moral Machine project

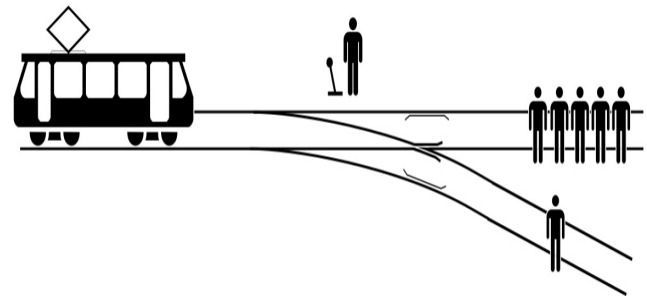
Ethics and Data

- Multiple areas for development of “data ethics”
 - **Data collection and handling** (e.g. generation, recording, curation, processing, dissemination, sharing)
 - **Data algorithms** (e.g. AI, artificial agents, machine learning, robots)
 - **Data practice** (e.g. responsible innovation, programming, hacking, professional codes, outcomes)



Utilitarian Ethics and the Trolley Problem

- **Utilitarian ethics:** The best action is the one that maximizes utility (e.g. well-being of humans)
 - ***What is good behavior?*** [Francis Hutcheson*]: Virtue is in proportion to the number of people a particular action brings happiness to. Vice / evil is proportionate to the number of people made to suffer.
- **Who should be saved (number → value)?**
 - Smaller vs. greater number of people?
 - You vs. others?
 - Your family vs. others?
 - “Good” people vs. “bad” people?



The Moral Machine Experiment – exploring utilitarian ethics for CAVs

- **Moral machine experiment:** Multilingual on-line serious game for collecting large-scale data on how citizens would want autonomous vehicles to solve moral dilemmas in the context of unavoidable accidents.
- **“Players”:** Individuals from 233 countries providing information about ~40M decisions
 - Geolocation information collected about users
 - Users could volunteer age, gender, income, education, religious and political attitudes
- **Play:** Users shown unavoidable accident scenarios with two possible outcomes, depending on whether vehicle swerves or stays on course. Users click on the outcome they find most preferable.

Participants and User Interface

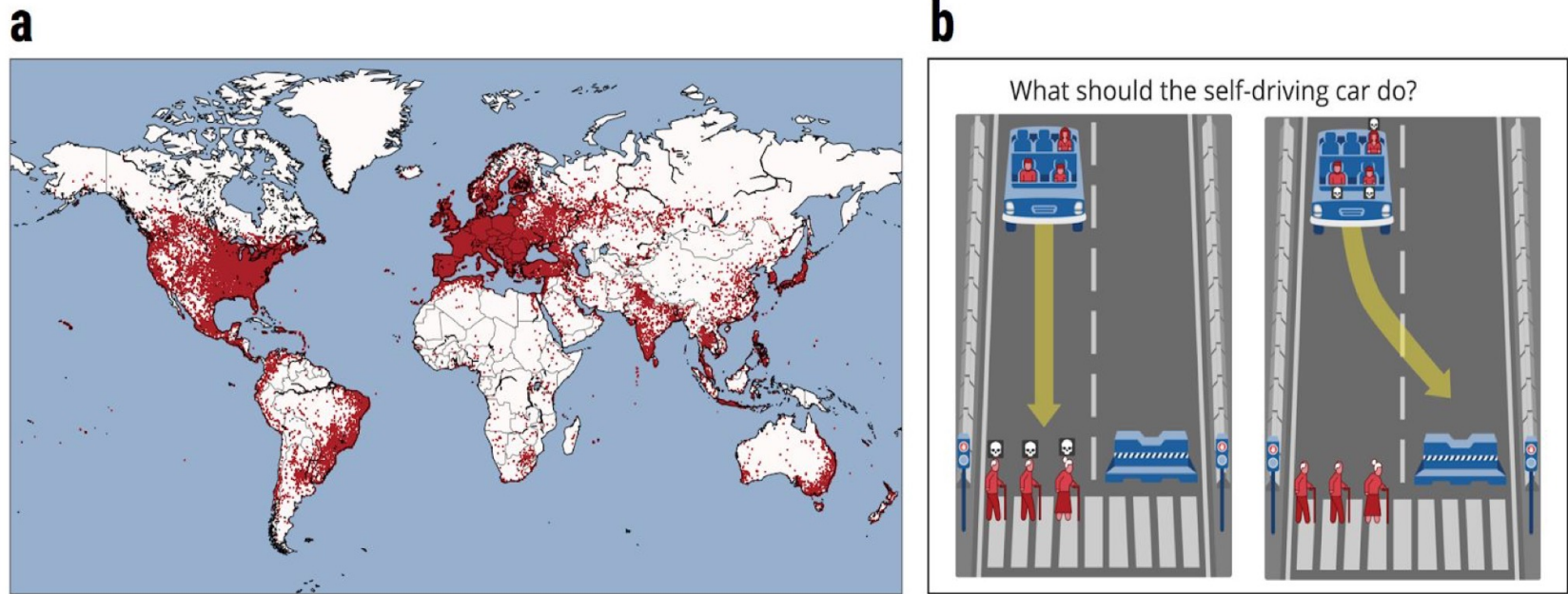


Figure 1. (a) World map highlighting the locations of Moral Machine visitors. Each point represents a location from which at least one visitor made at least one decision. The number of visitors or decisions from each location are not represented. **(b) Moral Machine interface.** An AV experiences a sudden brake failure. Staying on course would result in the death of two elderly men and an elderly woman, crossing on a “do not cross” signal (left). Swerving would result in the death of three passengers, an adult man, an adult woman, and a boy (right).

Ethical dilemmas

- Who should be spared?
 1. passengers vs. others
 2. humans vs. pets
 3. more lives vs. fewer lives
 4. men vs. women
 5. young vs. old
 6. pedestrians crossing legally vs. jaywalkers
 7. those who are fit vs. those who are less fit
 8. those with higher social status vs. those with lower social status
 9. Those who prefer to stay on course vs. those who prefer swerving (preference for action)

Experimental setup

- Deployed in June 2016
- In October 2016, feature added to offer users the option to fill a survey about their demographics, political views and religious beliefs
- 2016 – March 2017: website translated into 9 languages other than English (Arabic, Chinese, French, German, Japanese, Korean, Portuguese, Russian, Spanish)
- Country identified with user is geo-located from the IP address of the computer or mobile device.
- Source data (with anonymized IDs) made available to the community for replication, further research

Results

From <https://www.nature.com/articles/s41586-018-0637-6.pdf>

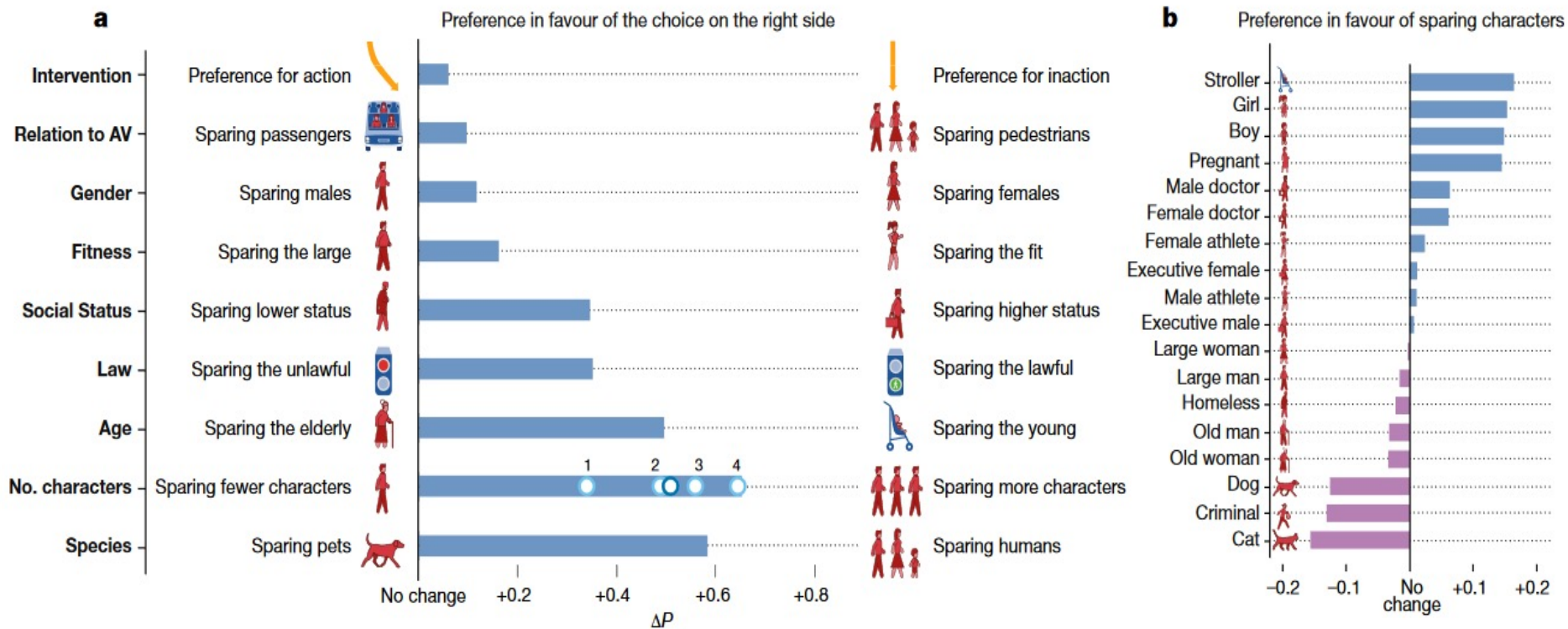


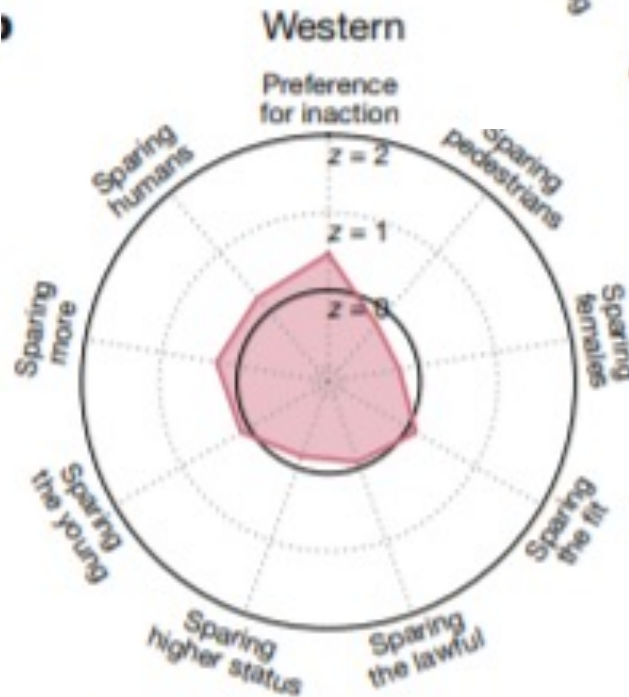
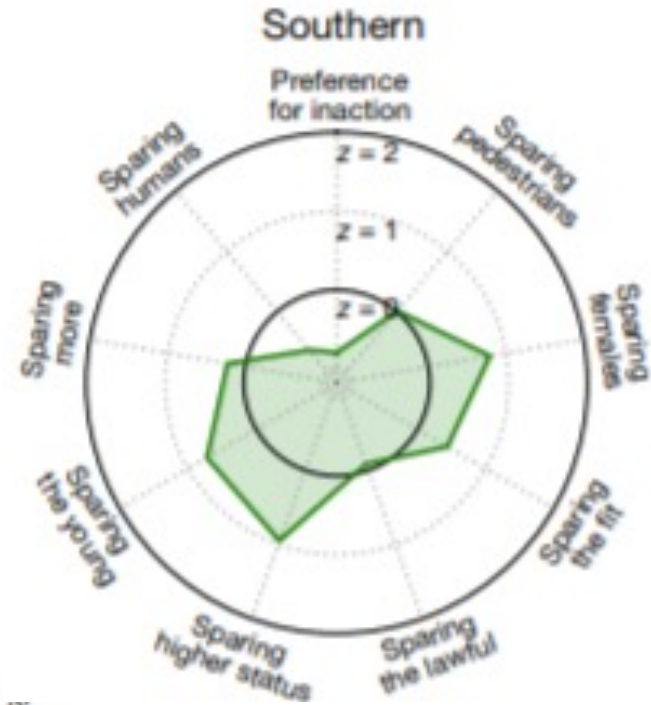
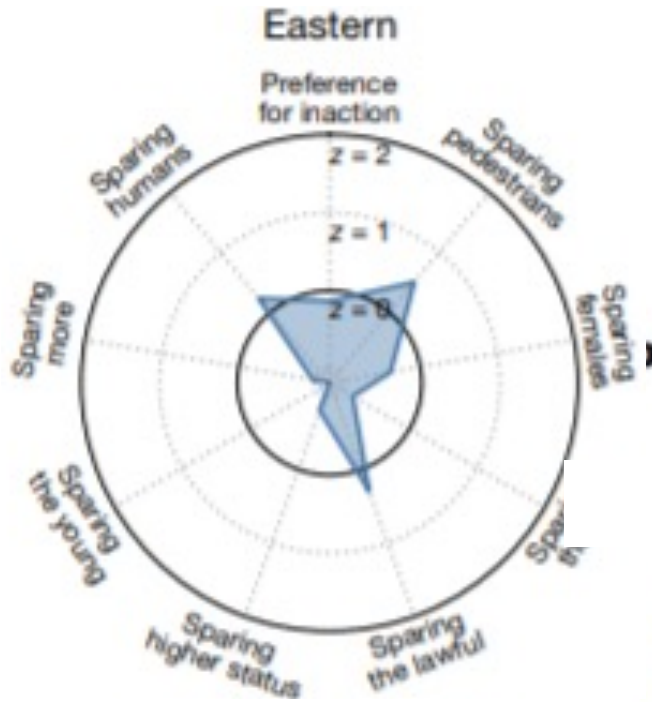
Fig. 2 | Global preferences. **a**, AMCE for each preference. In each row, ΔP is the difference between the probability of sparing characters possessing the attribute on the right, and the probability of sparing characters possessing the attribute on the left, aggregated over all other attributes. For example, for the attribute age, the probability of sparing young characters is 0.49 (s.e. = 0.0008) greater than the probability of sparing older characters. The 95% confidence intervals of the means are omitted owing to their insignificant width, given the sample size ($n = 35.2$ million). For the number of characters (No. characters), effect sizes are shown

for each number of additional characters (1 to 4; $n_1 = 1.52$ million, $n_2 = 1.52$ million, $n_3 = 1.52$ million, $n_4 = 1.53$ million); the effect size for two additional characters overlaps with the mean effect of the attribute. AV, autonomous vehicle. **b**, Relative advantage or penalty for each character, compared to an adult man or woman. For each character, ΔP is the difference between the probability of sparing this character (when presented alone) and the probability of sparing one adult man or woman ($n = 1$ million). For example, the probability of sparing a girl is 0.15 (s.e. = 0.003) higher than the probability of sparing an adult man or woman.

Data clustering

- Geolocation data used to identify 130 countries with >100 respondents.
- Clustering analysis identified 3 distinct “moral clusters” of countries:
 - **Western cluster:** North America and many European countries of Protestant, Catholic, and Orthodox Christian cultural groups. [Particular sub-clusters contained Scandinavian countries and another contained Commonwealth countries]
 - **Eastern Cluster:** Japan, Taiwan and other countries that belong to the Confucianist cultural group, as well as Islamic countries such as Indonesia, Pakistan and Saudi Arabia
 - **Southern Cluster:** Latin American countries of Central and South America. Also some countries characterized in part by French influence, France, French overseas territories and territories that were at some point under French leadership).

Moral Machine Clusters



Ethical predictors

Global preferences (strong preferences across all respondents)

- Spare humans over animals
- Spare more lives over fewer
- Spare young lives over old

Differences in ethical clusters (clusters differ in the weight they give to some preferences):

- **Individualistic cultures and collective cultures**
 - Respondents from **individualistic cultures** show greater preference for sparing more characters
 - Respondents from **collective cultures** show greater preference for sparing elders
- **Cultures with greater prosperity and quality of laws and institutions vs. those with lesser**
 - Greater preference for law-abiding pedestrians vs. jaywalkers
- Respondents from cultures with more **economic inequality** treat people of stature somewhat differently
- **Women were less expendable than men** in general *but* respondents in countries where there are higher ratios of female to male life expectancy and sex ratio at birth saw men as even less expendable than women

Caveats

- Study pool is big but **respondents were self-selecting**
 - Samples should not necessarily be construed as representative
- **No uncertainty** introduced about the classification of characters
 - In real life, one individual may have many attributes and varying value to the respondent
- **No hypothetical relationships** between respondents and characters assumed (family, friends, etc.)
- Study **not translated into every language**, only way to take it was **online**

ARTICLE

<https://doi.org/10.1038/s41586-018-0637-6>

The Moral Machine experiment

Edmond Awad¹, Sohan Dsouza¹, Richard Kim¹, Jonathan Schulz², Joseph Henrich², Azim Shariff^{3*}, Jean-François Bonnefon^{4*} & Iyad Rahwan^{1,5*}

With the rapid development of artificial intelligence have come concerns about how machines will make moral decisions, and the major challenge of quantifying societal expectations about the ethical principles that should guide machine behaviour. To address this challenge, we deployed the Moral Machine, an online experimental platform designed to explore the moral dilemmas faced by autonomous vehicles. This platform gathered 40 million decisions in ten languages from millions of people in 233 countries and territories. Here we describe the results of this experiment. First, we summarize global moral preferences. Second, we document individual variations in preferences, based on respondents' demographics. Third, we report cross-cultural ethical variation, and uncover three major clusters of countries. Fourth, we show that these differences correlate with modern institutions and deep cultural traits. We discuss how these preferences can contribute to developing global, socially acceptable principles for machine ethics. All data used in this article are publicly available.

We are entering an age in which machines are tasked not only to promote well-being and minimize harm, but also to distribute the well-being they create, and the harm they cannot eliminate. Distribution of well-being and harm inevitably creates tradeoffs, whose resolution falls in the moral domain^{1,2}. Think of an autonomous vehicle that is about to crash, and cannot find a trajectory that would save everyone. Should it swerve onto one jaywalking teenager to spare its three elderly passengers? Even in the more common instances in which harm is not inevitable, but just possible, autonomous vehicles will need to decide how to divide up the risk of harm between the different stakeholders on the road. Car manufacturers and policymakers are currently struggling with these moral dilemmas, in large part because they cannot be solved by any simple normative ethical principles such as Asimov's laws of robotics³.

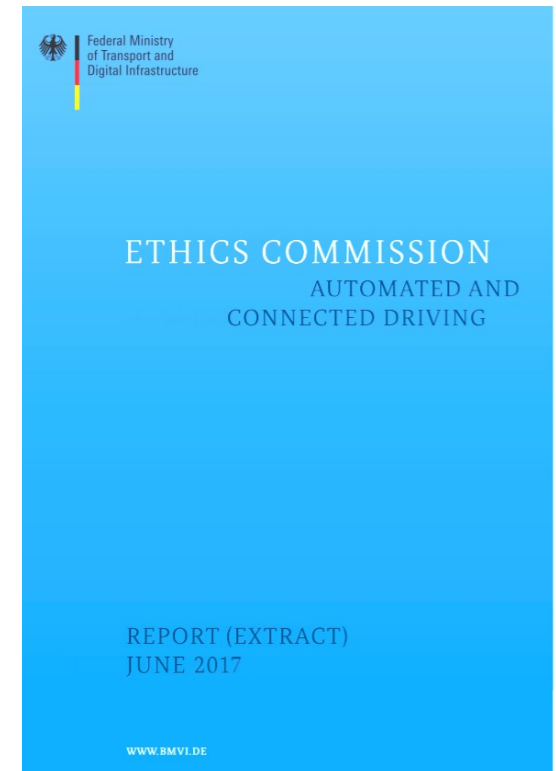
Asimov's laws were not designed to solve the problem of universal machine ethics, and they were not even designed to let machines

vehicles, and for the wider public to accept the proliferation of artificial intelligence-driven vehicles on their roads, both groups will need to understand the origins of the ethical principles that are programmed into these vehicles⁴. In other words, even if ethicists were to agree on how autonomous vehicles should solve moral dilemmas, their work would be useless if citizens were to disagree with their solution, and thus opt out of the future that autonomous vehicles promise in lieu of the status quo. Any attempt to devise artificial intelligence ethics must be at least cognizant of public morality.

Accordingly, we need to gauge social expectations about how autonomous vehicles should solve moral dilemmas. This enterprise, however, is not without challenges⁵. The first challenge comes from the high dimensionality of the problem. In a typical survey, one may test whether people prefer to spare many lives rather than few^{6,12,13}, or whether people prefer to spare the young rather than the elderly^{14,15}, or whether people prefer to spare pedestrians who cross legally, rather

An initial pass at CAV ethics ...from the German Federal Ministry of Transport and Infrastructure

- Ethics Commission of the Federal Ministry of Transport and Digital Infrastructure in Germany released ethical guidelines for self-driving cars in 2016.
- Commission included 14 scientists and legal experts, and Ministry said it would implement and enforce the guidelines.
- Germany home to major automakers such as BMW, Daimler and Volkswagen.



https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?__blob=publicationFile

2016 Ethical rules for automated and vehicular traffic 1 (excerpt, paraphrased)

1. Primary purpose of autonomous vehicles is to improve safety and increase mobility.
2. Prevent accidents when possible
3. Save people over everything else
4. Public sector responsible for guaranteeing safety through regulation, policy, enforcement
5. Liability or damage should be governed by the usual product liability principles
6. It should be clear who is controlling the car. Drivers should receive training in the operation of autonomous vehicles.

2016 Ethical rules for automated and vehicular traffic 2 (excerpt, paraphrased)

7. In accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited
8. **Genuine dilemmatic decisions, such as the decision between one human life and another cannot be clearly standardized, nor can they be programmed**
9. Permitted business models must respect limitations in the autonomy and data sovereignty of road users
10. Complete connectivity and central control of all motor vehicles is ethically questionable if it is unable to safely rule out the total surveillance of road users and manipulation of vehicle control.

How should ethical decisions be made for CAVs?

- If you had to make an ethical decision while driving, would you choose based on the high-level categories in the Moral Machine experiment??
- Could characteristic data (law abiding vs. law breaking, high status vs. lower status) used in the Moral Machine experiment be assessed by an automated system?
- How should CAVs make these decisions?

Lecture 19 Sources (not already on slides)

- “How to Build Ethical Algorithms”, The Conversation,
<http://theconversation.com/ethics-by-numbers-how-to-build-machine-learning-that-cares-85399>
- “What is Data Ethics?”,
<http://rsta.royalsocietypublishing.org/content/374/2083/20160360>
- The Moral Machine Experiment, <https://www.nature.com/articles/s41586-018-0637-6.pdf>
- The Moral Machine Experiment: 40 Million Decisions and the Path to Universal Machine Ethics,
<https://dspace.mit.edu/bitstream/handle/1721.1/125065/Moral%20Machine%20Paper.pdf;jsessionid=4B9B1E98968EB874764CD3628DEB943D?sequence=1>

Presentations



Upcoming Presentations

April 15

- **“New wave of hacktivism adds twist to cybersecurity woes,”** Reuters, <https://www.reuters.com/article/uk-cyber-hacktivism-focus/new-wave-of-hacktivism-adds-twist-to-cybersecurity-woes-idINKBN2BH3I3>
- **“Clop ransomware gang breaches University of Colorado and University of Miami,”** Security Magazine, <https://www.securitymagazine.com/articles/94891-clop-ransomware-gang-breaches-university-of-colorado-and-university-of-miami>

April 19

- **“This cuffing season, it’s time to consider the privacy of dating apps”**, Brookings Institution, <https://www.brookings.edu/blog/techtank/2020/11/20/this-cuffing-season-its-time-to-consider-the-privacy-of-dating-apps/>
- **“How private is your on-line dating data?”**, Consumer Reports, <https://www.consumerreports.org/privacy/how-private-is-your-online-dating-data/> [NEED VOLUNTEER]

April 22

- **“What a gambling app knows about you”**, New York Times, <https://www.nytimes.com/2021/03/24/technology/gambling-apps-tracking-sky-bet.html>
- **“Can computer algorithms learn to fight wars ethically?”**, Washington Post, https://www.washingtonpost.com/magazine/2021/02/17/pentagon-funds-killer-robots-but-ethics-are-under-debate/?no_nav=true&tid=a_classic-iphone

Today's Presentations

April 11

- **“Vaccine passports pose ethical thicket for Biden Administration,”** Politico, <https://www.politico.com/news/2021/03/17/vaccine-passports-ethics-biden-administration-476384> [Davis]
- **“This is bigger than just Timnit’: How Google tried to silence a critic and ignited a movement”**. Fast Company, <https://www.fastcompany.com/90608471/timnit-gebru-google-ai-ethics-equitable-tech-movement> [Grant]