

CMSC 304 Computer Ethics

Dr. Cynthia Matuszek

1

Bookkeeping

- EA4 is out
- Today: self-driving cars
- Please get your exam from me if you have not

2

Where are we on self driving cars?

- Level 0 (No Driving Automation): There may be systems in place to help the driver.
- Level 1 (Driver Assistance): A single automated system for driver assistance, such as steering or accelerating (cruise control).
- Level 2 (Partial Driving Automation): This means **advanced driver assistance systems** or **ADAS**. The vehicle can control both steering and accelerating/decelerating. A human sits in the driver's seat and can take control of the car at any time.
- Level 3 (Conditional Driving Automation): Environmental detection capabilities and can make informed decisions for themselves but still require human override.
- Level 4 (High Driving Automation): Can intervene if things go wrong or there is a system failure. Do not require human interaction *in most circumstances*.
- Level 5 (Full Driving Automation): Do not require human attention.

3

The Technology of the Car

- Anti-Lock Brakes
- Electronic Stability control
- Adaptive cruise control
- Lane-departure warning system
- Self parking
- Automated guided vehicle systems
- Automated braking (~15% reduction in accidents)
- Collision warning systems (~7% reduction in accidents)

4

Moore's Law of self-driving cars

- GM imagined self-driving cars by 2070 in the 50s
- Metric: The number of miles auto-driven between **disengagements**
 - In which the safety driver had to intervene unexpectedly
- This number doubles every 16 months, very very roughly
 - Depends on the company, what's considered a disengagement, publicly reported numbers, etc.
- Per Toyota as of late 2019, miles between disengagement will reach human performance in 2039
 - Until then we're stuck with ADAS

5

Human monitoring

- Existing ADAS systems, and everything before level 5, requires human monitoring and preparedness to take over
- Say such a disengagement is necessary every 11,000 miles
 - Waymo's publicly reported value c. 2019
- Will people really pay attention for 10,000 miles of smooth operation?
- Danger zone between expectation and reality

6

Tesla

- Probably the current leader in distribution of ADA systems
 - Despite the Autopilot and "Full Self-Driving" names
- 4 formal investigations into safety complaints since 2018
- 15 recalls since January of 2021
- At least 33 crashes and 11 deaths
- "Selected Tesla drivers have been beta testing the "Full Self-Driving" software on public roads... drivers must be ready to intervene at any time"

https://www.breitbart.com/tech/2021/04/01/tesla-autopilot-safety-complaints/

7

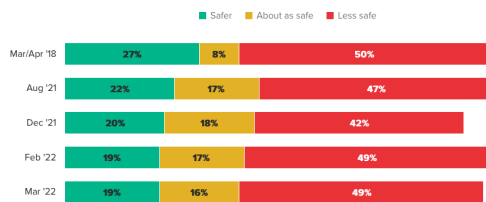
Current state of NON-autonomous vehicles

- Approximately 40,000 drivers/passengers die each year in auto accidents
 - 1.18 fatalities for every 100 million miles traveled.
 - Americans drive nearly 3.2 trillion miles each year.
- About 6,000 pedestrians die each year in automobile accidents
- 90%+ of accidents can be attributed to human error
- If avoided, could save hundreds of billions in time and cost and loss of life

Slide adapted from Thomas P. Barnett, Commons Conference 2011

8

Public perception



How safe are self-driving cars compared to human drivers?

https://www.pewresearch.org/2022/04/06/autonomous-vehicles-safety-and-use/

9

Perception vs Reality

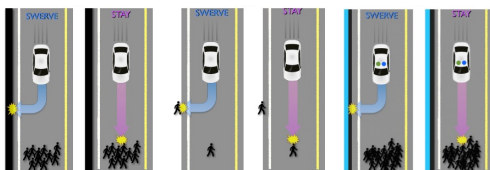
- 73% of people are afraid to ride in a fully-autonomous vehicle.
- Only 40% of people believe autonomous vehicles are safer than regular cars.
 - Went down after Uber accident when it was 50% (Arizona, 2018)
- Over 80% people believe they are an above average safe driver.**
- Up to 60% of drivers would not choose an AV if offered the option.
- Most drivers will sacrifice themselves to save others, but almost all drivers will sacrifice strangers before family.
 - Utilitarian ... up to a point.

Slide adapted from Thomas P. Barnett, Commons Conference 2011

10

Ethical dilemmas

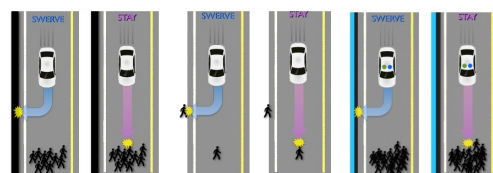
- At heart, the problem is that cars can hurt or kill people.
- Assuming some damage is unavoidable, who should get hurt?
- Can be presented as an **ethical dilemma**



Slide adapted from Thomas P. Barnett, Commons Conference 2011

11

How Does Ethical Dilemma Survey Work?



- Would you sacrifice yourself for a small group or would you hit the small group?
- Would you swerve to kill someone or let the vehicle hit the person in the cross-walk?
- Would you swerve to sacrifice yourself for the large group?

Slide adapted from Thomas P. Barnett, Commons Conference 2011

12

The (in)famous trolley problem

- A trolley's brakes have failed.
- You are controlling the signal switch.
- If you do nothing, five people will be killed.
- If you activate the switch, only one person will be killed.
- What do you choose to do?
- Critical distinction: **Allowing** death versus **causing** death

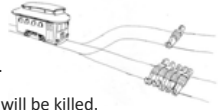


image: sarahzurro.medium.com/the-trolley-problem-73e22048d89e
Slide adapted from cammus.univie.ac.at/academic/cultivewide/58f2019/Team1/Slides.pdf

13

The bridge variant

- You are on a bridge watching a runaway trolley
- It will soon hit 5 people
- You can stop it by pushing someone from the footbridge to its path
- If you do nothing, 5 people will be killed
- If you push the person, 1 person will be killed
- Same distinction, but **causing** death more directly

14

The basic problems

1. This never happens
 2. These decisions are not generally clear as to outcomes
 3. Generally discussing probabilities and distributions
 4. Systems generally do not have clear picture of environment
- "Trolley Problems are constructed in such a way where the agent has one choice, and knows with certainty what will happen should she make that choice. Moreover, that choice is constructed as a dilemma: it appears that no matter what choice she makes, she will end up doing some form of wrong."

Source: https://www.researchgate.net/publication/312204889

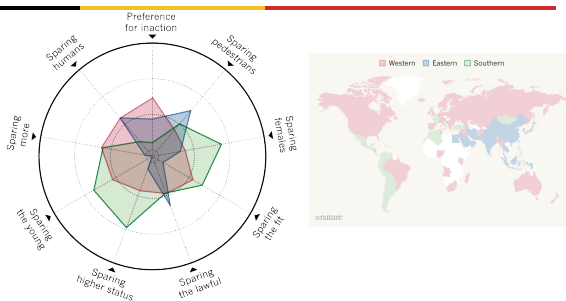
15

The Moral Machine

- The ultimate dilemma-based SD car experiment
 - Young or old, rich or poor, more people or fewer?
- There are no universal ethical rules
 - Varies heavily by country, culture, and background
- People want an autonomous vehicle to protect pedestrians even if it meant sacrificing its passengers — but also won't buy self-driving vehicles programmed to act this way
- Clustering gives 3 large groups

16

The Moral Machine



17

Some additional ethical questions

- Privacy and Security: Who will own the data that is collected?
- Explainability and Transparency: Will the different data types be cleared explained and how these data types can be used so end users understand the risks and implications?
- Accountability: Who is the ultimate owner/of the engineering underpinnings of the driverless car - the automobile manufacturer?
- Data Bias: Who will set the rules governing accident decisions on who will be killed with certainty or be spared ?
- Sustainability: as there is high energy consumption requirements moving to driverless cars, what are the environmental risks vs other driving methods?

18

A new class of victims

- There will be an inherent shift in the makeup of automobile accident victims.
- E.g., decrease in driver deaths, increase in pedestrian and cyclist deaths.
- Great news for some people, bad news for others.
- Can we accept an increase in the death rate of certain groups of people if it means a decrease in the overall death rate?

Note: selected from common sense, not necessarily evidence-based. (http://tiny.cc/9d-cars)

19

Public acceptance of AVs

- How many (more) lives per year must be saved for the public to embrace a driverless revolution?
- Manufacturers will need to be upfront with the details of the decision-making systems piloting the vehicles
- Will there be override capabilities?
- Delayed Feedback Problem: There will be a significant period of time before we've collected enough data to determine the effectiveness of a driverless initiative

Note: selected from common sense, not necessarily evidence-based. (http://tiny.cc/9d-cars)

20

The really important question

- How good do self-driving cars need to be before we deploy them widely?
- How much better than humans?
- Roughly 1.18 fatalities for every 100 million miles driven in the United States in 2016
- Need several hundred million miles to understand if an AV is a better driver than humans, on average
 - We're probably in the low tens of millions now

21

Another non-obvious question

- Who's responsible for negative outcomes for AVs?
- The manufacturer?
- The human driver?
 - If it's explicitly an ADAS? If it's explicitly level 4 or 5?
- The person who purchased the car?
- The programmers?
- The machine learning model?
 - We are not generally comfortable without an entity to blame

22

Discussion

<http://tiny.cc/sd-cars>

- **When fully autonomous self-driving cars are better than humans, should human driving be banned? Why or why not?**
 - <1 fatality per 100 million miles
- How should self-driving cars be tested and developed?
 - On the road with real people?
 - The way Tesla is doing it?
 - On closed tracks, even if it slows down development?
- Who should ultimately have responsibility for negative outcomes of self-driving vehicles?

23