Week 10 Notes

Constantine Kaganis, Ethan Fiddle, Richard Payne

The Moral Machine Experiment

- Collected data from various regions and made charts to show probabilities of choosing one side over another in a situation where an ethical decision is made, in the context of autonomous vehicle accidents
- Collected data based off of culture clusters
  - Western, Eastern, Southern
  - Eastern part cares mores about following the law, less about sparing the young, do not follow Bentham model
  - Should be taken into account by regulators so the laws that are formed are modeled based on the preferences of that country

The Moral Machine Experiment (continued)

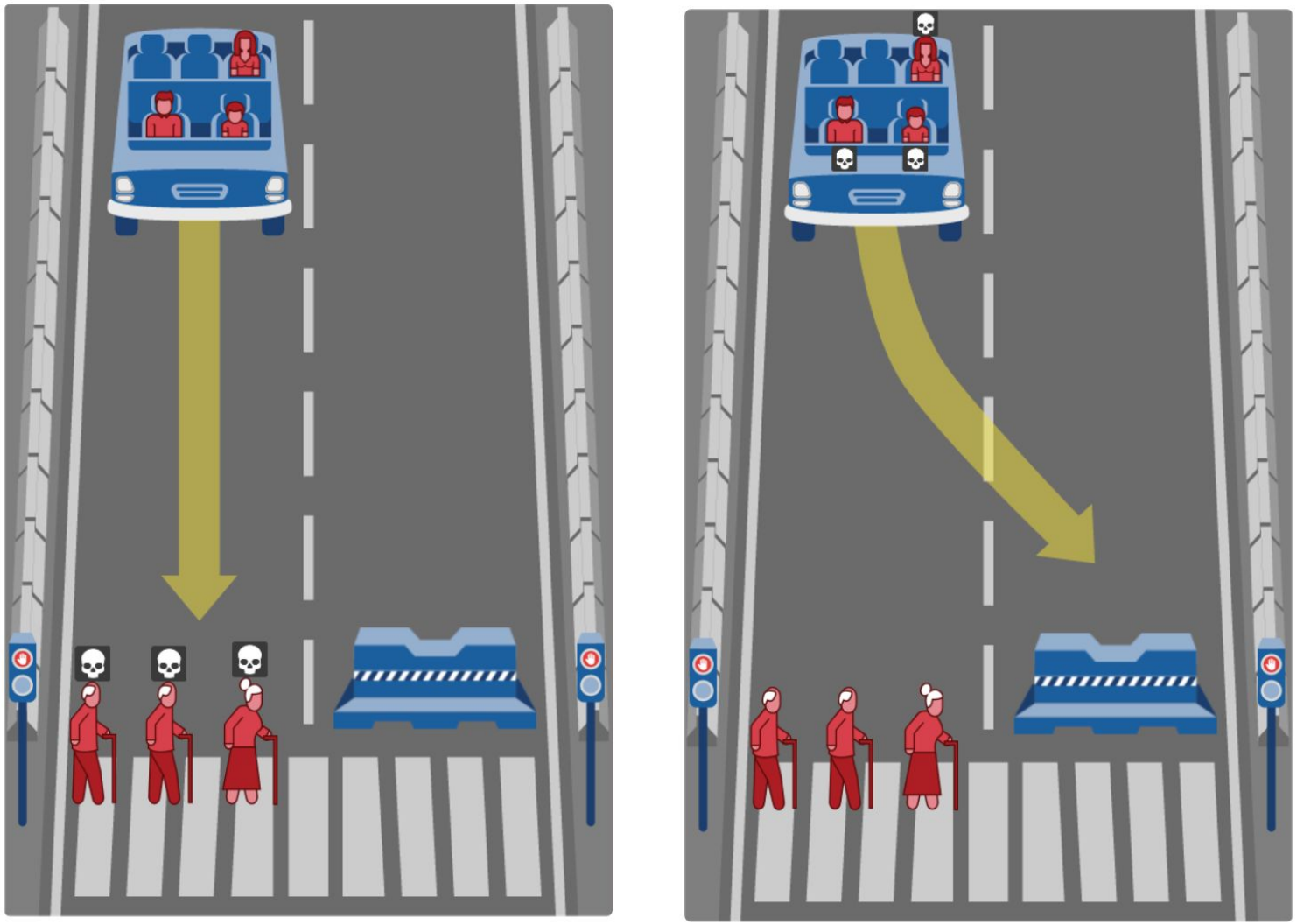
- 2016 autonomous vehicle study gathered around 40 million decisions in 10 different languages from many people spanning across 233 countries
  - Moral preferences among those who participated in the study:
    - Young lives prioritized
    - Human lives valued more over animals
  - Spared most often were:
    - babies in strollers
    - pregnant women
    - Children
  - Final results also showed ethics varying between different cultures

The Moral Machine Experiment (continued)

- Let the car take its course. Even if it will harm more people. Do not take an action that will harm others.
- If the man pulls the lever, he makes the conscious choice and kills a person. If he doesn’t pull the lever, he had nothing to do with the trains original path.
- Human beings are not rational. We are controlled by emotions.
- People themselves would want a car that protects themselves but wants everyone else to buy a car that will protect others.

The Moral Machine Experiment (continued)

What should the self-driving car do?



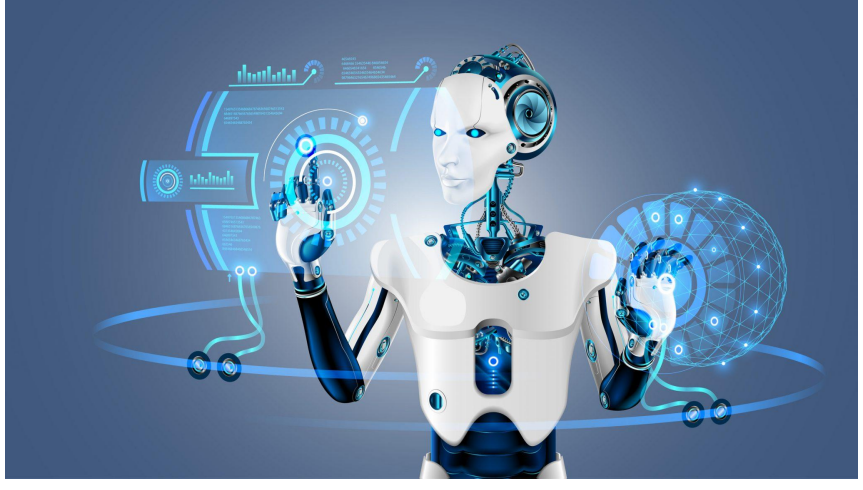
Iyad Rahwan’s Ted Talk

- Talk centered around technology and society, and self-driving cars
  - social aspects of technology
- Moral decision making
- Bentham line of thought: minimize death count
- Dilemma: do you take action to minimize the death count? Or do not take action and let the situation play out
- Social dilemma: if everyone thinks, “I won't follow the rules, but I hope everyone else will”, then no one will follow the rules
  - This is why collecting user data on ethical preferences is important
- Asimov’s laws of robotics

Iyad Rahwan's Ted Talk - Isaac Asimov's Law of Robotics

- Isaac Asimov's 3 Laws of Robotics:

1. A robot may not injure a human being or allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings except if they were to conflict with the first law.
3. A robot must protect its own existence - may not allow itself to come to harm



Prisoner's Dilemma

- Originally created by Merrill Flood and Melvin Dresher while working at RAND
  - RAND (Research and Development)
    - RAND is an American nonprofit created in 1948
- A situation in which two players both have two options, whose outcome depends on the choices made by the other
- Rational decision is to always defect for your best interest, in that the risk reward is greatest for this decision where you either go free or you get 1 year
- Lost Angeles
- Models many phenomena

THE PRISONER'S DILEMMA

	B stays silent (cooperates)	B betrays A (defects)
A stays silent (cooperates)	Both serve 1 year	A serves 3 years, B goes free
A betrays B (defects)	A goes free, B serves 3 years	Both serve 2 years

Prisoner’s Dilemma (continued)

- Each player’s choice: Cooperate or Defect
- If both cooperate, one chocolate each
- If both defect, no chocolates for anyone
- If one defect, one cooperate: Defect gets two chocolates, cooperate owes me a chocolate

- Defect is the safest option. You either get nothing or gain 2
- If you cooperate you either get 1 or lose 1.

Nuclear Arms Rivalry

- US vs USSR (1950s)
- Build H-bomb or not
- Each nation prefers: build h-bomb and the other does not
- Little gained if both build h-bomb (money wasted)
- Game theory predicts: both build h-bomb

Game Theory: Normal Form Games

- A normal form game is a type of game which includes all possible strategies, as well as their payoffs, for each player within that game
- Prisoner's Dilemma is an example of a normal form game
  - John von Neumann - Hungarian-American mathematician and computer scientist

- List of players, strategies, payoffs
- Simultaneous
- Zero-sum
- Rock paper scissors

Nash Equilibrium

- Set of strategies for each player within a game, such that no player has any reason to change his or her strategy based on how the other players are playing
- Can show that decision makers follow randomized strategies