



THE UNIVERSITY OF  
MELBOURNE

# Ethics in Artificial Intelligence

---

Associate Professor Tim Miller  
School of Computing and Information Systems  
University of Melbourne  
[tmiller@unimelb.edu.au](mailto:tmiller@unimelb.edu.au)  
[@tmiller\\_unimelb](https://twitter.com/tmiller_unimelb)

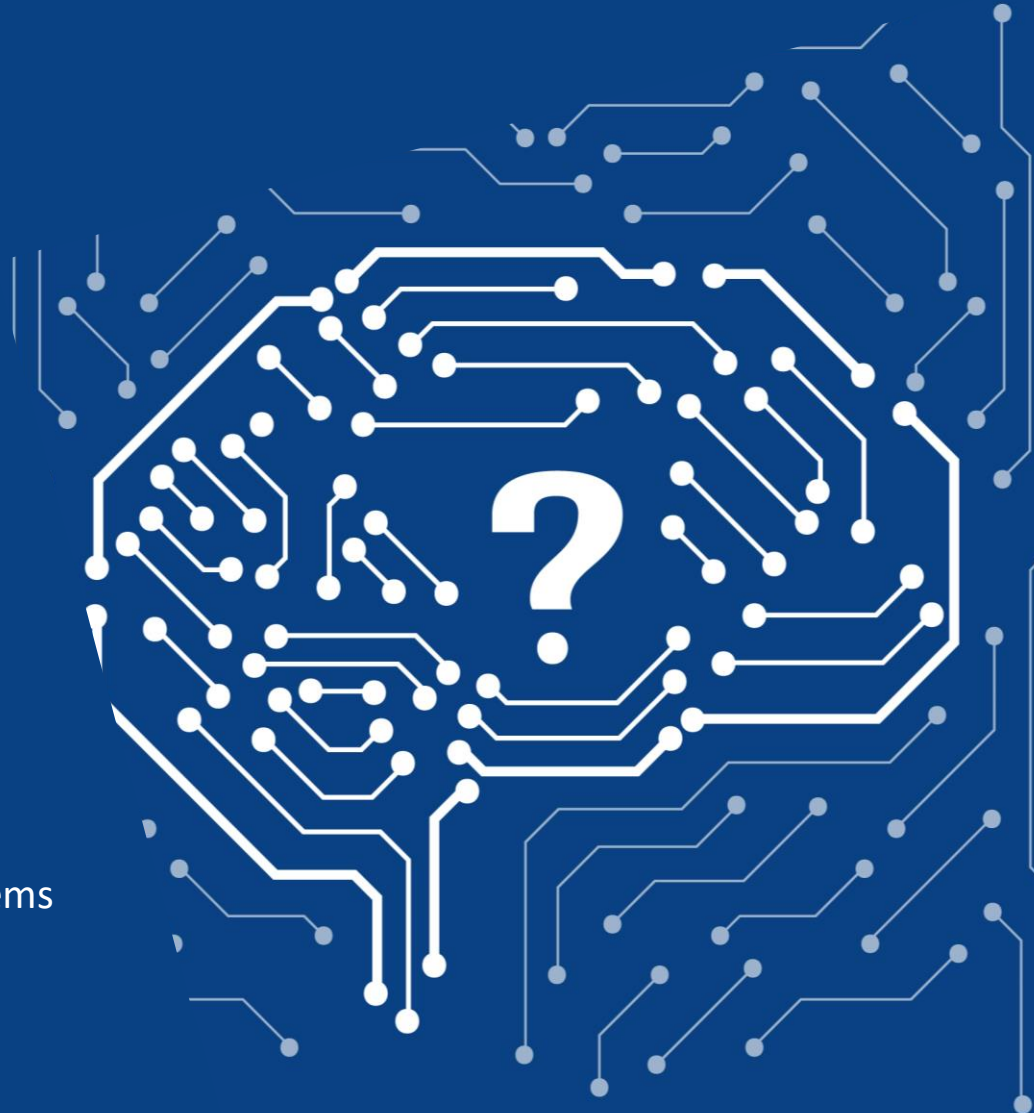




Image source: <https://www.google.com/selfdrivingcar/>

# The Trolley Problem

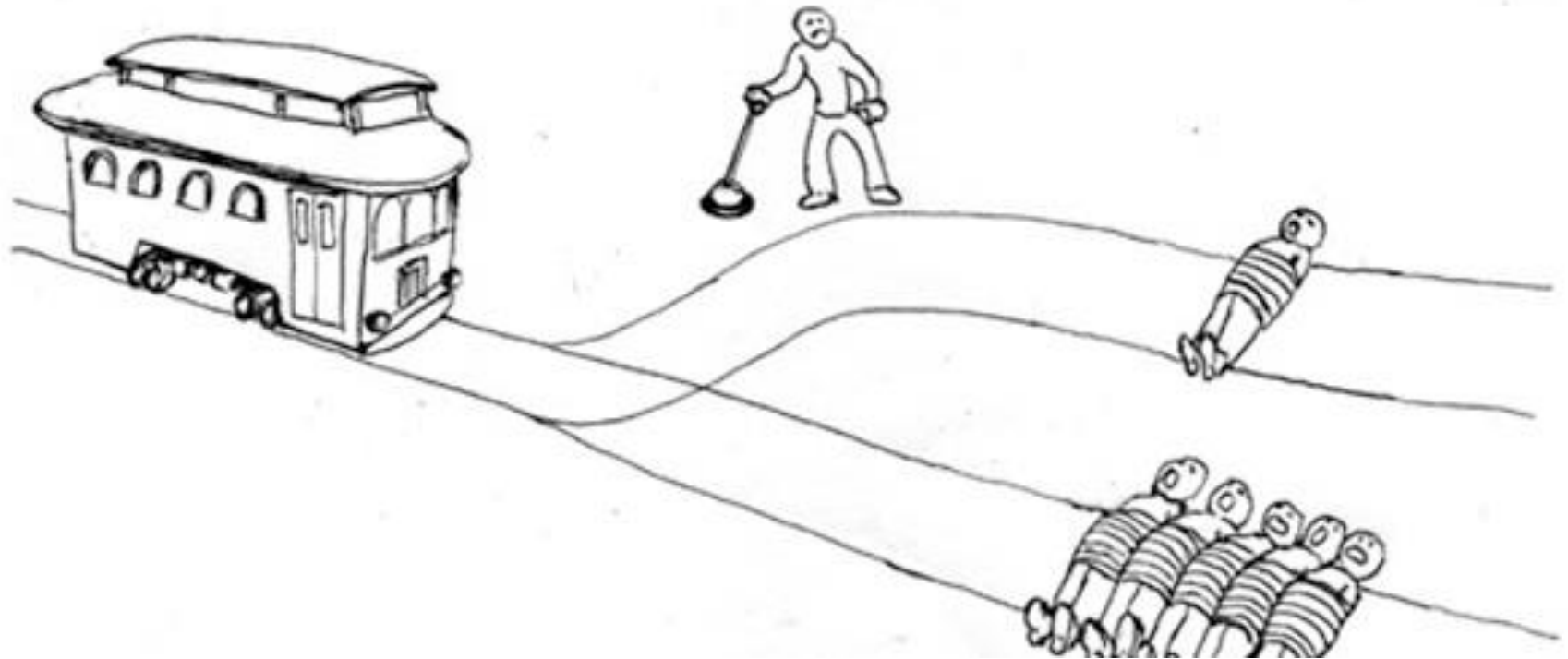


Image source: <http://www.relativelyinteresting.com/the-trolley-problem-a-thought-experiment-that-tests-our-morality/>



# Ethics

*Utilitarianistic ethics:* only outcomes matter, and we should always opt for the best.

*Deontological ethics:* actions matter too! We have a duty to help others as well as a duty not to harm others, but the latter is stronger than the former.

*What should we do?*



# The 'Fat Man' Trolley Problem

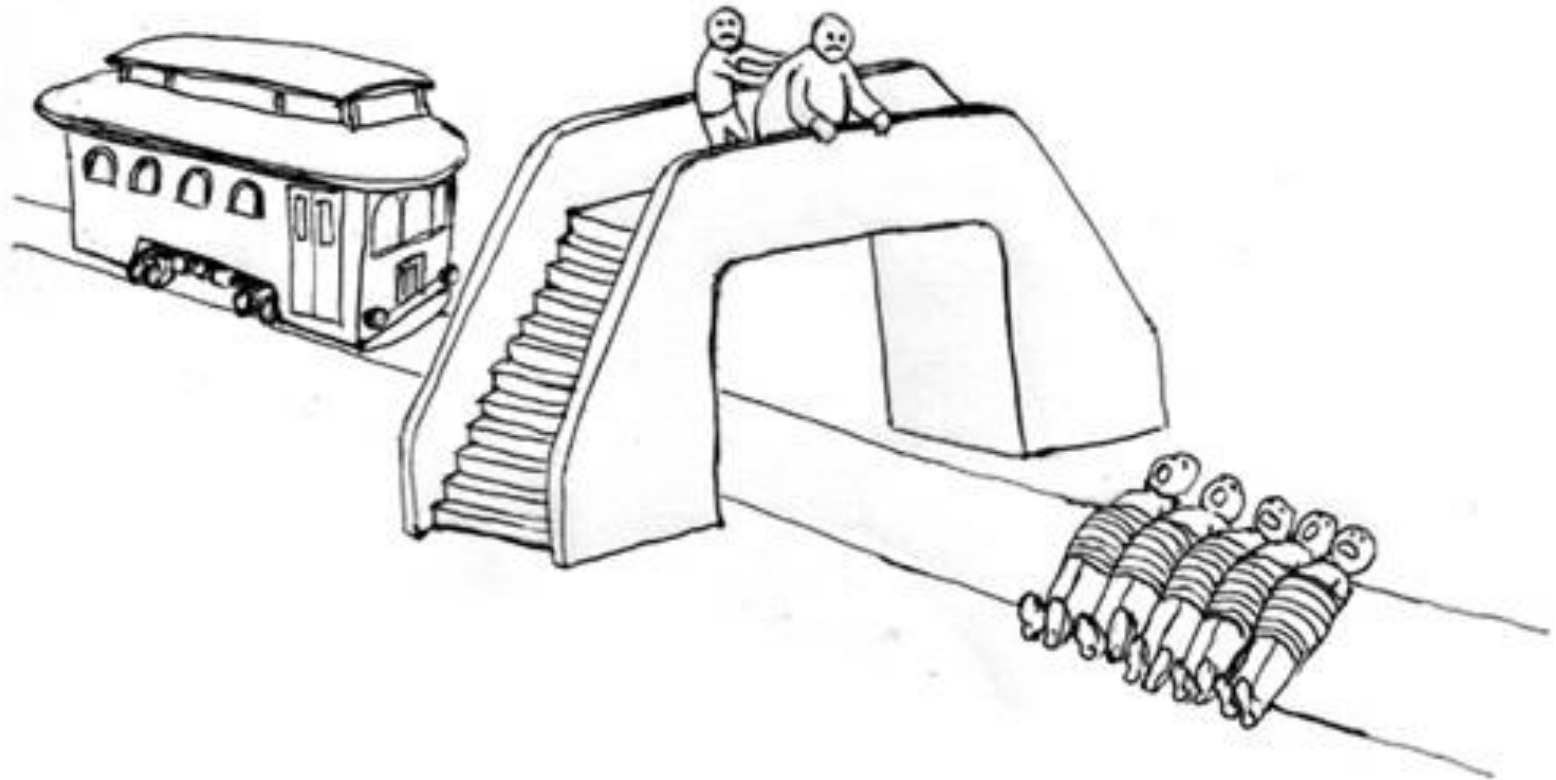


Image source: <http://www.relativelyinteresting.com/the-trolley-problem-a-thought-experiment-that-tests-our-morality/>



# Many variants of the Trolley Problem

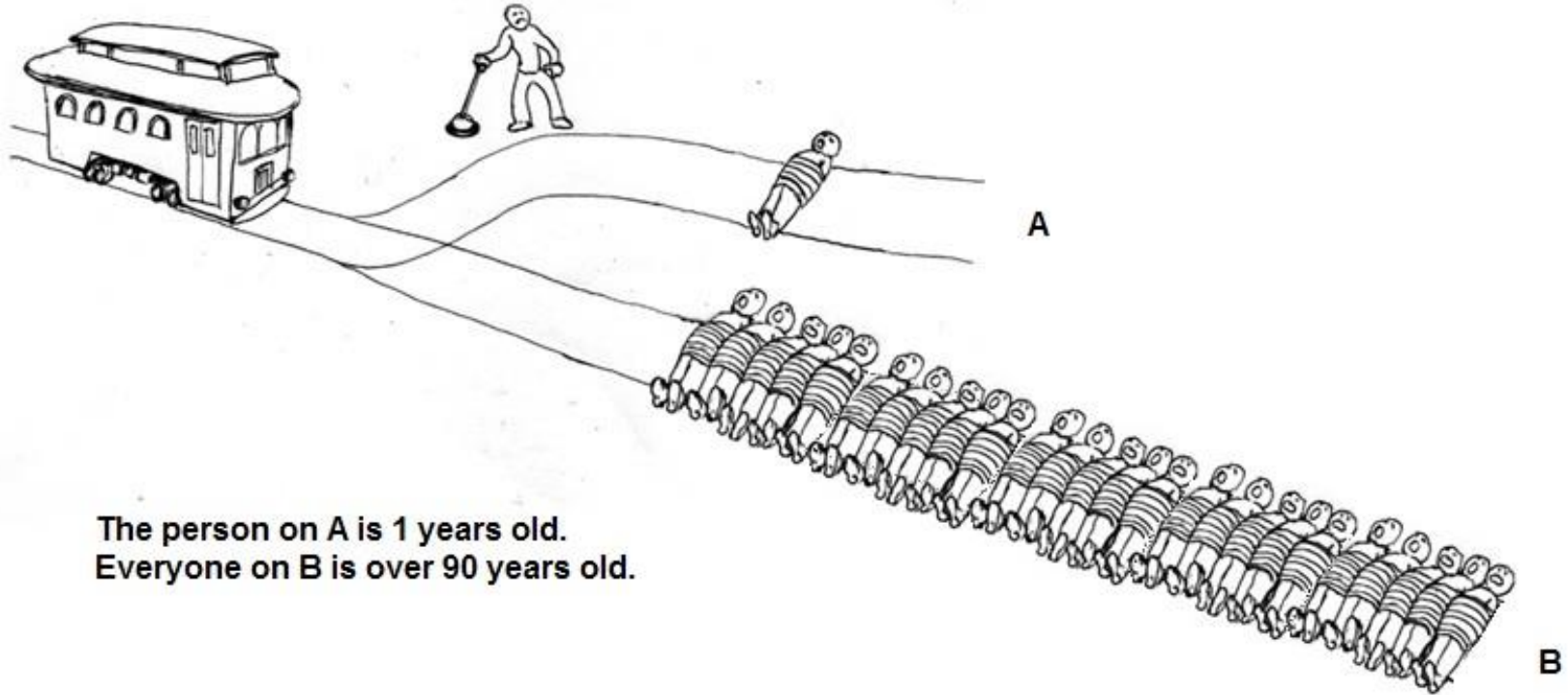
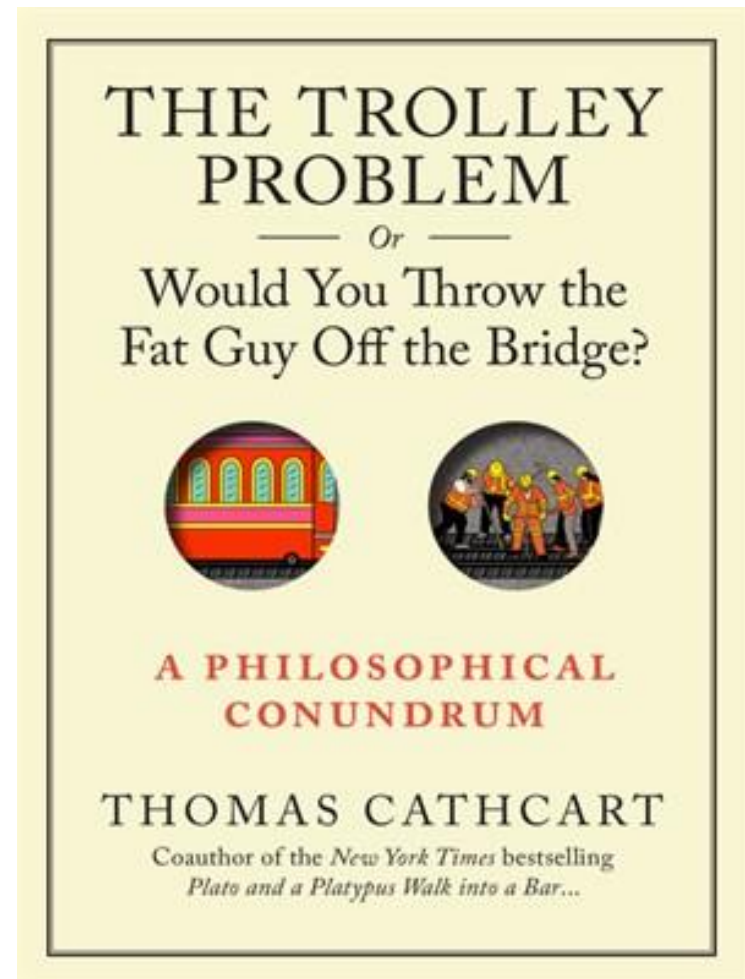


Image source: <http://www.relativelyinteresting.com/the-trolley-problem-a-thought-experiment-that-tests-our-morality/>

# Experimental ethics





# Pre-programmed ethics

Split-second driver decision:

- reckless driver → criminal liability
- otherwise → blame “fate”, “bad luck”, “heat of the moment”, and do not blame the driver

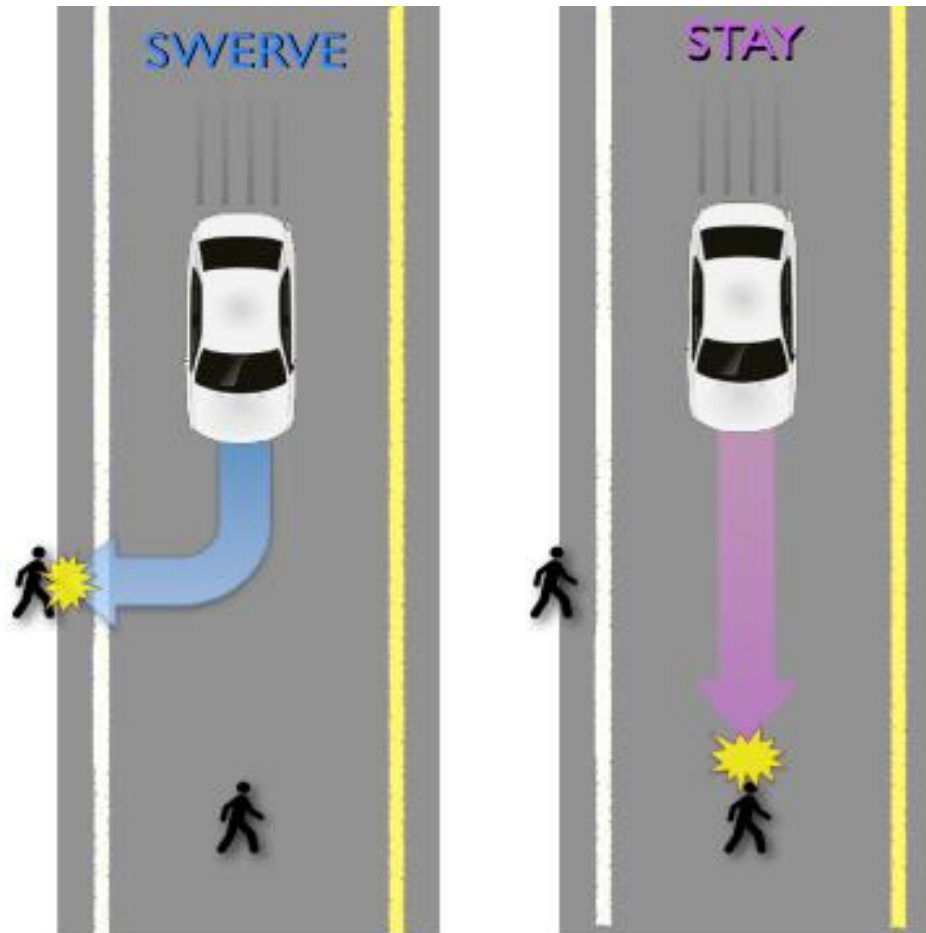
Automated decision-making:

- We must decide in advance! No split-second decisions
- What behaviour should be programmed?

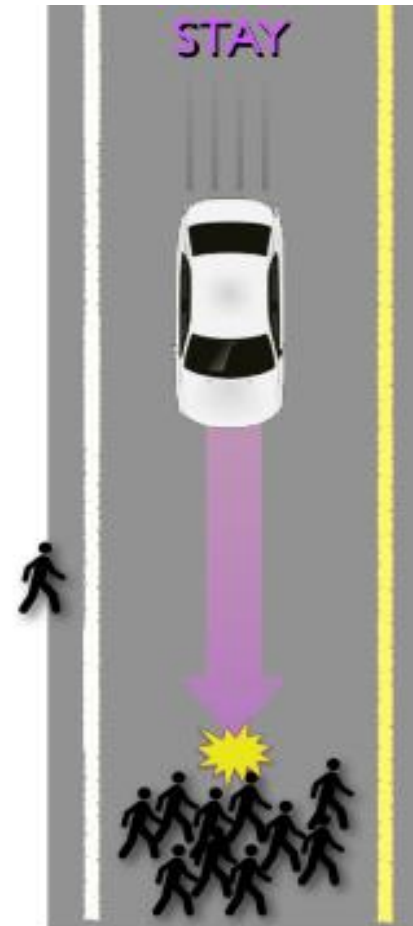
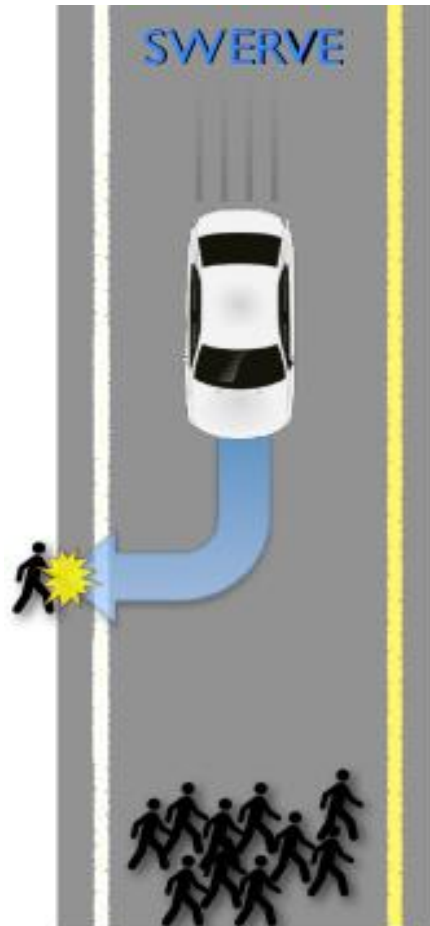


# Swerve or stay?

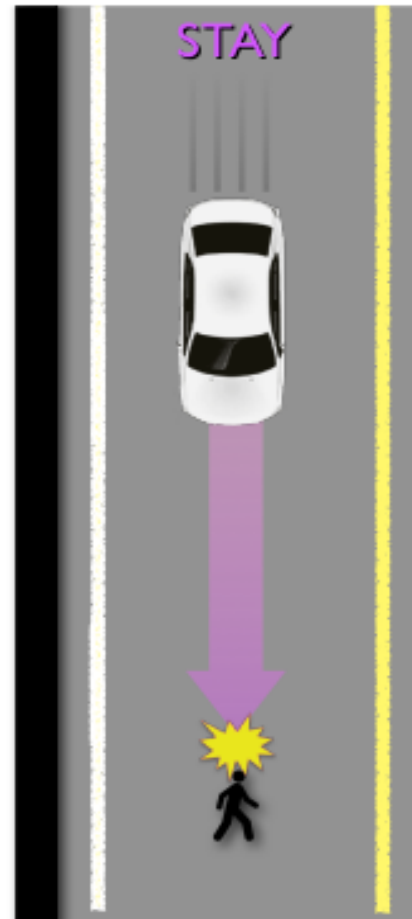
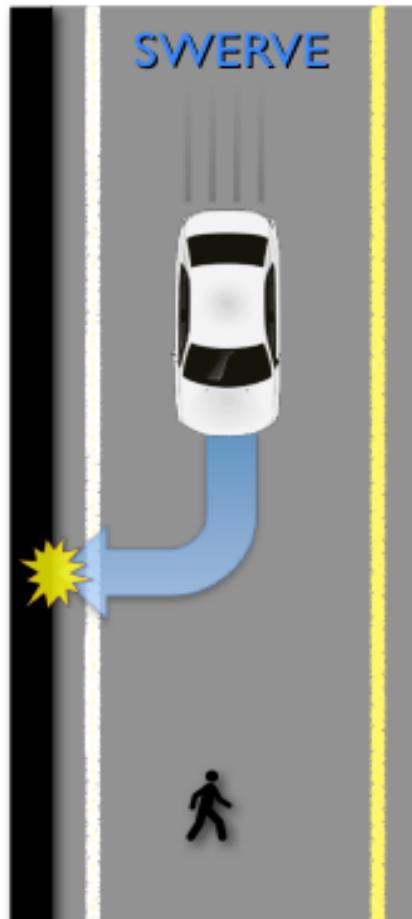
<http://go.unimelb.edu.au/i4rr>



# Swerve or stay?



# Swerve or stay?





# In some larger-scale experiments

Would you buy a self-driving car that protected passengers instead of the driver?

Despite preferring self-sacrificing vehicles ...

... respondents were less likely to buy such cars.

Trade-offs must be made by law-makers to solve this *social dilemma*



# But.....

What about related but different scenarios:

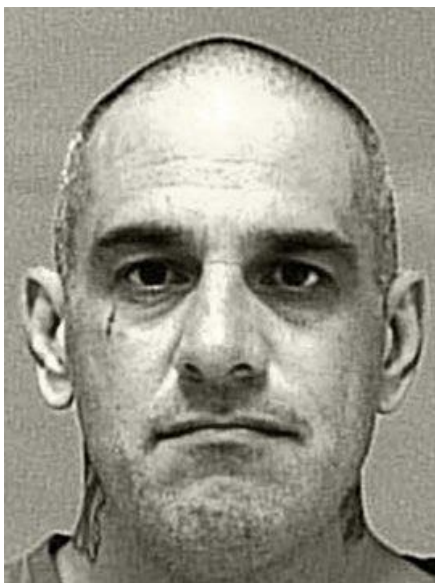
- 99.95% confidence an object is not a person
- What if the cars knows there is a child on board?
- Buyer liability: If buyers can choose the “moral algorithm” are they liable for its choices?

## Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

*by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica*

May 23, 2016



### **Vernon Prater**

#### **Prior Offenses:**

- 2 armed robberies
- 1 attempted armed robbery

**Rating: Low (3)**

### **Brisha Borden**

#### **Prior Offenses:**

- 4 juvenile misdemeanours

**Rating: High (8)**





# Amazon drops secret AI recruiting tool that showed bias against women

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process



▲ Amazon's automated hiring tool was found to be inadequate after penalizing the résumés of female candidates.  
Photograph: Brian Snyder/Reuters



# Who should decide?

AI programmers?

Companies?

Lawmakers?

The public?

Ethicists and philosophers?





# The Moral Machine at MIT Media Lab

<http://moralmachine.mit.edu/>



THE UNIVERSITY OF  
MELBOURNE

# Thank you

---

tmiller@unimelb.edu.au  
@tmiller\_unimelb

