

# Ethics in Artificial Intelligence

Associate Professor Tim Miller  
School of Computing and Information Systems  
University of Melbourne



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

# The Trolley Problem

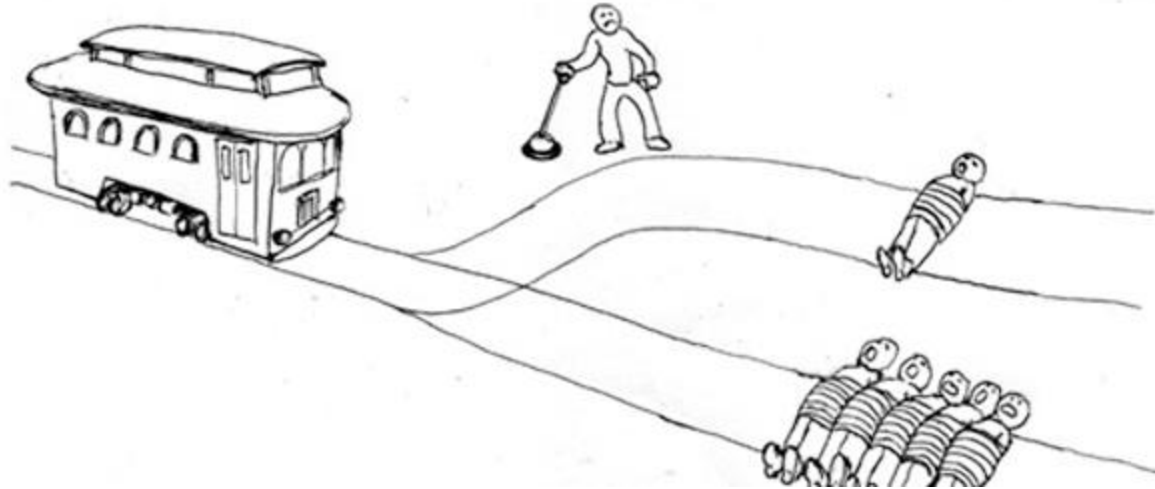


Image source: <http://www.relativelyinteresting.com/the-trolley-problem-a-thought-experiment-that-tests-our-morality/>

# 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

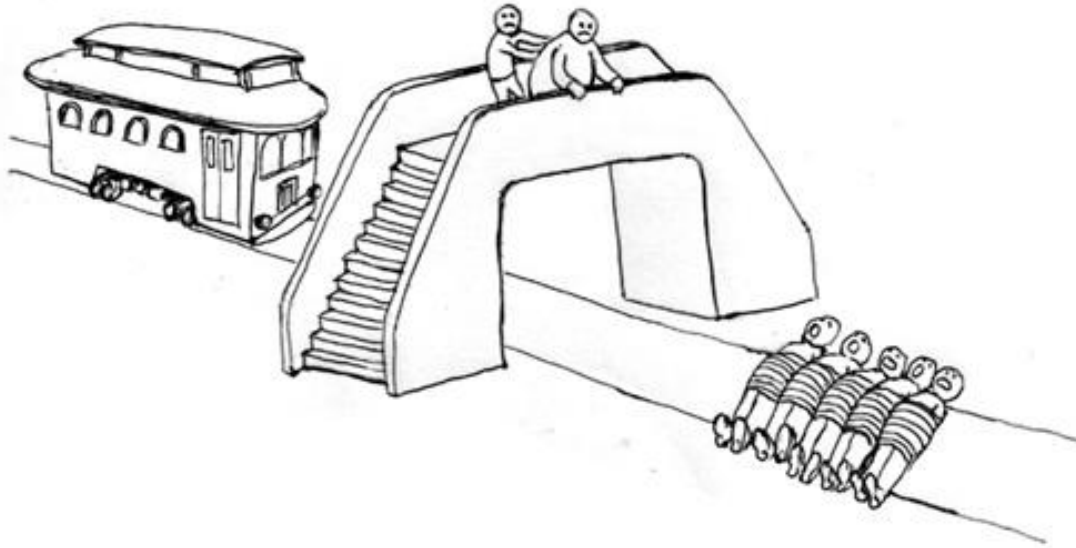


Image source: <http://www.relativelyinteresting.com/the-trolley-problem-a-thought-experiment-that-tests-our-morality/>

# Many variants of the Trolley Problem

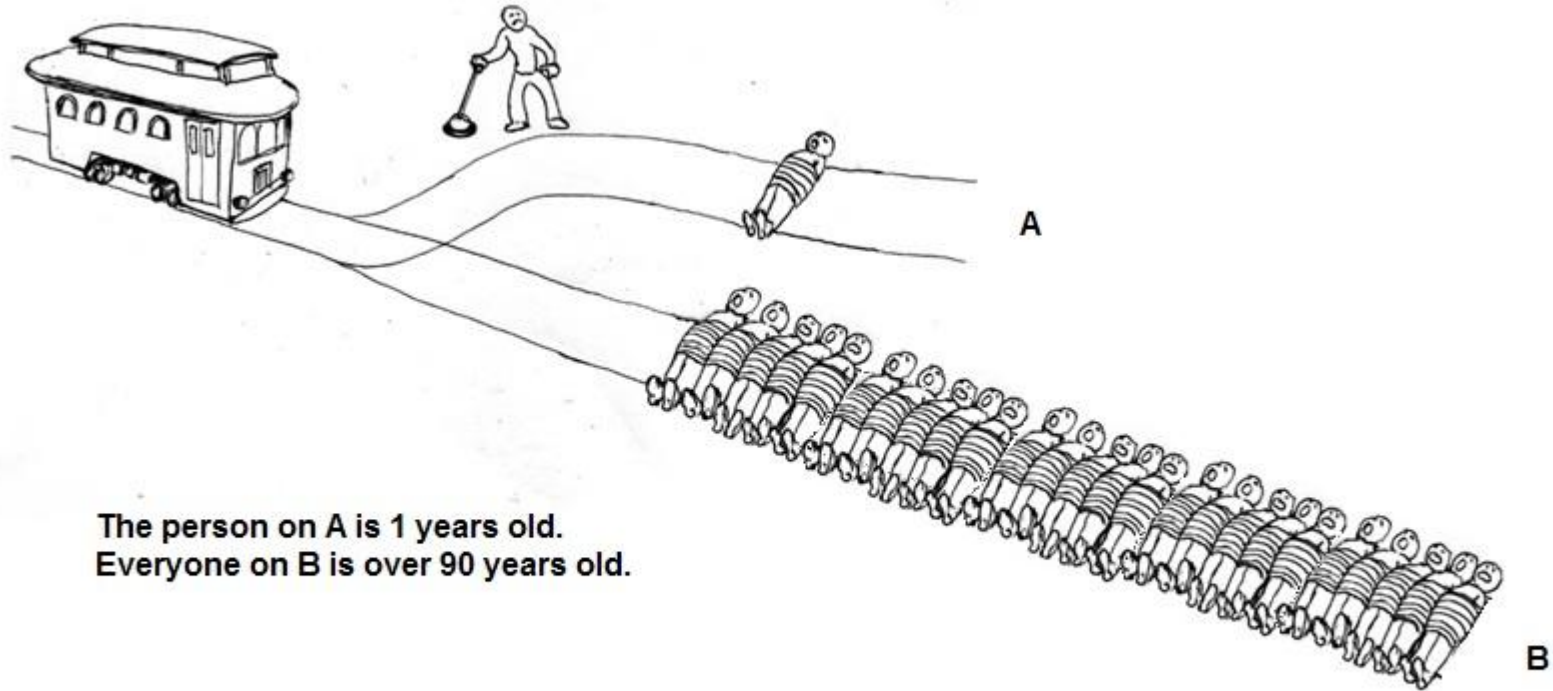
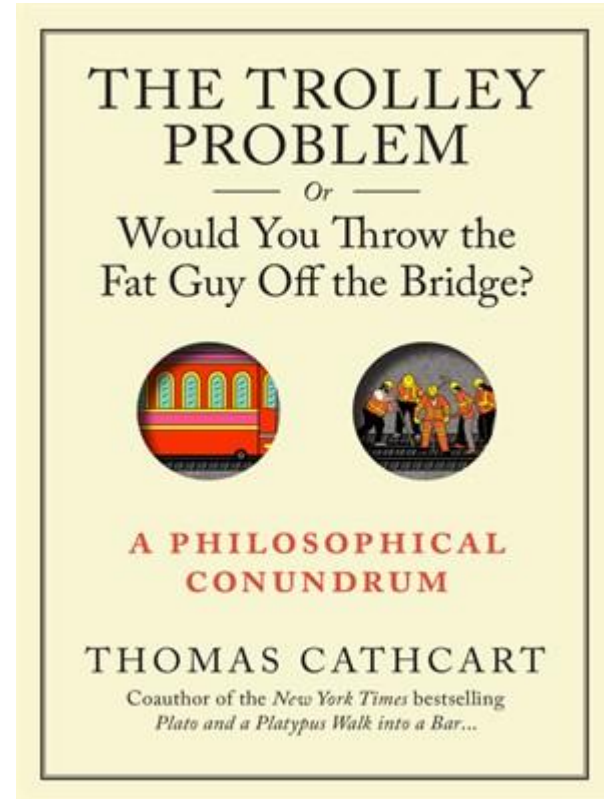
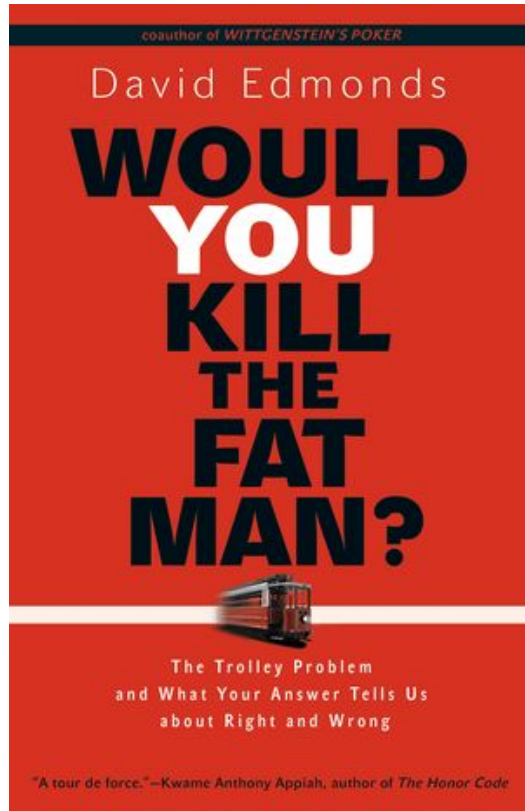


Image source: <http://www.relativelyinteresting.com/the-trolley-problem-a-thought-experiment-that-tests-our-morality/>

# Experimental ethics



# Pre-programmed ethics

Split-second driver decision:

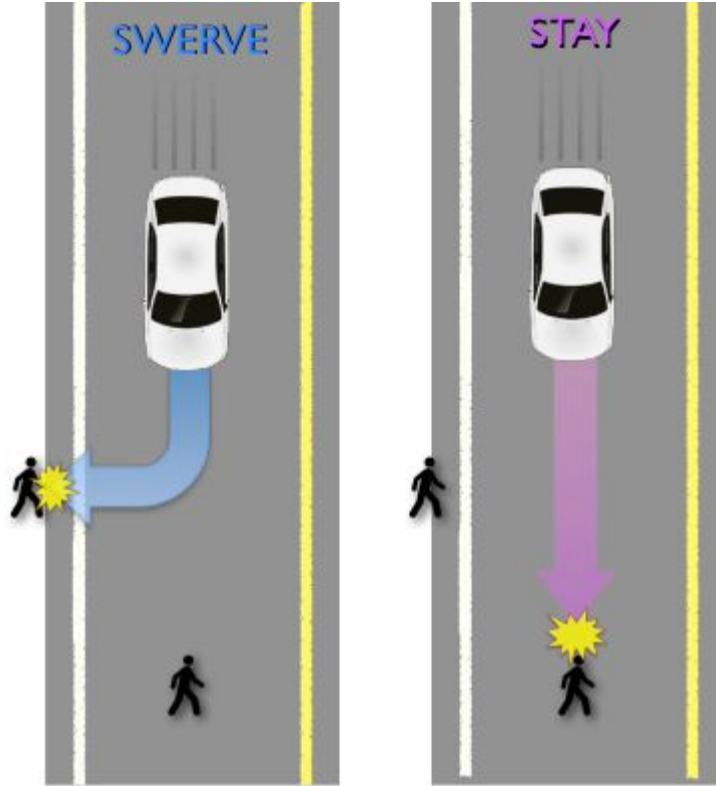
- “reckless driver” → criminal liability
- otherwise → blame “fate”, “bad luck”, “heat of the moment”, and do not blame the driver

Autonomous vehicles:

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



# Swerve or stay?



<https://www.surveymonkey.com/r/COMP90054>

# Pedestrian vs. Passenger

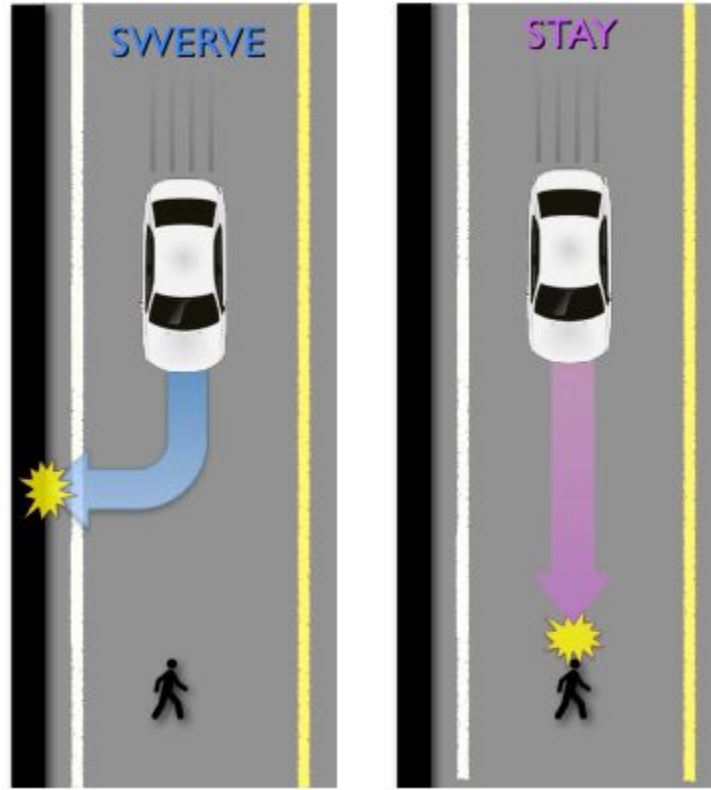


Image sources: [Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?](#) JF Bonnefon, A Shariff, I Rahwan, arXiv preprint arXiv:1510.03346

# In some larger-scale experiments

Despite preferring self-sacrificing vehicles ...

people were less likely to indicate willingness to buy such cars.

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

# But.....

What about slightly different scenarios:

- 20% chance of killing 10 pedestrians vs 80% chance of killing 1
- 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?

# Who should decide the behavior of self-driving cars?

AI programmers?

Car manufacturers?

Lawmakers?

The public?

Psychologists and philosophers???

# The Moral Machine at MIT Media Lab

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

# Wifi Access

1. Ensure **Wi-Fi** is enabled on your device.
2. Select **Visitor** from available wireless networks.
3. Launch a **web browser** and access any website (On some devices this is done automatically). Your web browser will redirect to the **Visitor login screen**.
4. Enter the **Visitor username** and **password** *very carefully*:
  - a. Username: [sciencediscoveryday](#)
  - b. Password: [Toa@Zm](#)
5. Click **Connect/Ok**.