Consequentialism

Giovanni Sartor

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

The concept of consequentialism

- An action is morally required
  - iff it delivers that best outcome, relative to its alternative
  - Iff its good outcomes outweigh its negative outcomes to the largest extent
  - Iff it produces the highest utility?

- Morality as an optimisation problem!

- Various kinds of consequentialism
  - What are the good and bad things to be maximised?
  - How many there are?
  - How much each of them matters?
  - Can we construct a single utility function that combines gains and losses over multiple valuable goals?

The reference approach: Utilitarianism

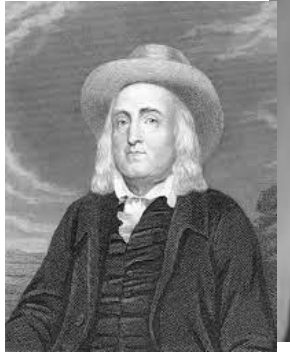
- Jeremy Bentham,
- John Stuart Mill. From Utilitarianism (1861). Principle of utility:
  - Actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure

- Utility: Happiness or satisfaction of desires/interests
- Utilitarianism is not egoism
  - The utility of everybody has to be taken into account equally

Advantages of utilitarianism

- Conceptually simple
- Egalitarian (everybody's utility counts in the same way)
- Fits with some basic intuitions (making people happy is good, making them suffer is bad)
- In many case it is workable, in some cases problematic (what should we do about hunger, how shall we treat friends and relatives, etc.)

Two versions of utilitarianism




- Act utilitarianism
  - Do the action that maximises utility
  - Do the optifimic action
- Rule utilitarianism
  - Follow the rule the consistent application of which maximises utility
  - Follow the optifimic rule

- Is AI utilitarian
  - What utility function would be utilitarian?
  - Should AI systems adopt an utilitarian reward function?
  - Should they go for the Act or the Rule versions (are they Archangels or Proles?)

Issues with act utilitarianism

- Does it provide a good decision procedure
  - Can we choose what to do by optimising the outcome our actions? Do we have the information to make this calculation? Can an AI system have the information?
- Does is provide a good standard for assessing decisions?

- What is the link between utility and a reward function?

Act utilitarianism: Problems

- Is it too demanding.
  - Should I give to the poor all that I above the minimum that allows me to survive?
  - Should I give the same importance to everybody, regardless of their connection to me?
  - Is it OK to harm some people for the greater benefit of others
    - Reprisals? Torture? Sadism?

- What could an utilitarian say:
  - The cases in which utilitarianism seems to fail are not realistic
  - There is no real contrast between utilitarianism and mainstream moral beliefs

Rule utilitarianism

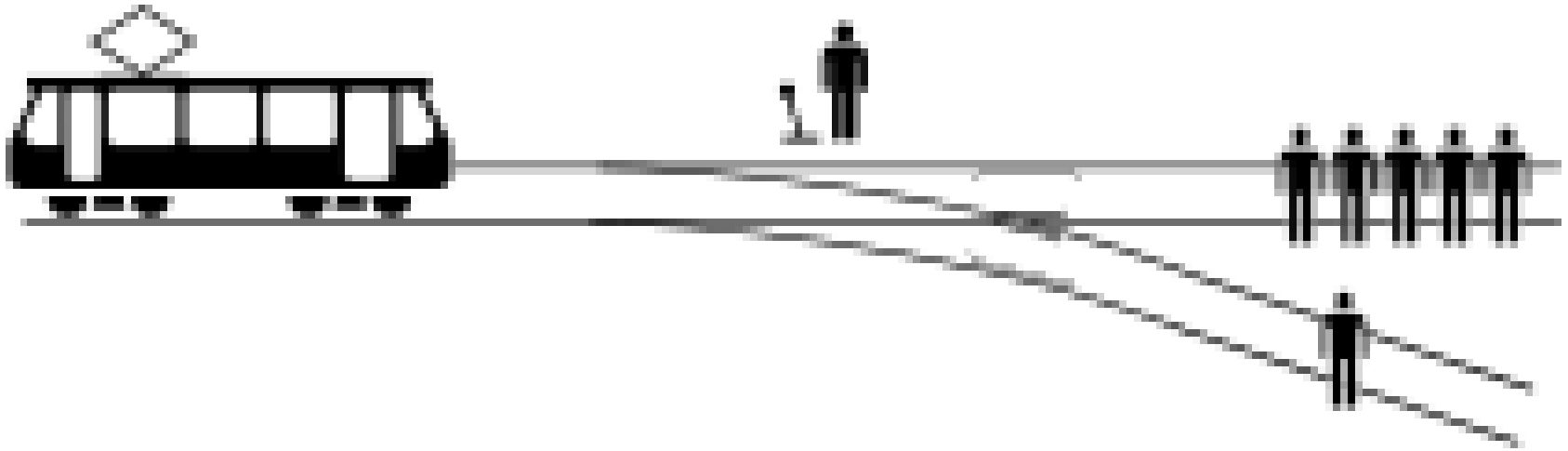
- an action is morally right just because it is required by an optimific social rule (a social rule the general compliance with which would provide the highest utility)
  - It is ok to tell the truth, not to steal, etc. since the general compliance with such norms would deliver the greatest utility
  - What about those exceptional cases in which the rule does not deliver
  - What is you know that most people are not following the rule.
    - Should we be honest if most people around as are dishonest?

A further issue: distribution

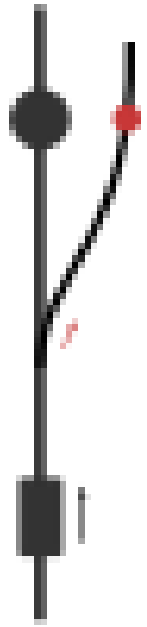
- Does it matter how the good and bad outcomes are distributed?
  - It is ok to make an action that benefits some to the detriment of others?
  - Always if the benefits outweigh disadvantages?

- Utilitarianism vs wealth maximisation
  - Utilitarianism favours (modest) redistribution of wealth, since the same amount of money gives more utility to the poor than to the reach
  - The impact of redistribution on wealth generation however has to be considered
- Wealth maximisation (adopted by some economic approach) aims at maximising the wealth in society regardless of distribution.

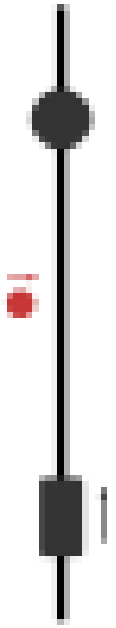
The trolley problem



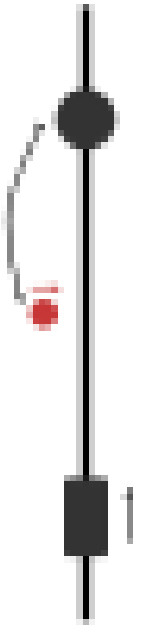
What would you do? What should an AI system tasked with monitoring traffic do



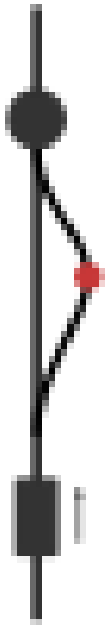
the switch
Ford, 1967



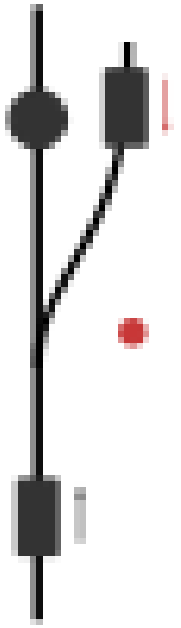
the fat man
Thomson, 1978



the tri-villain

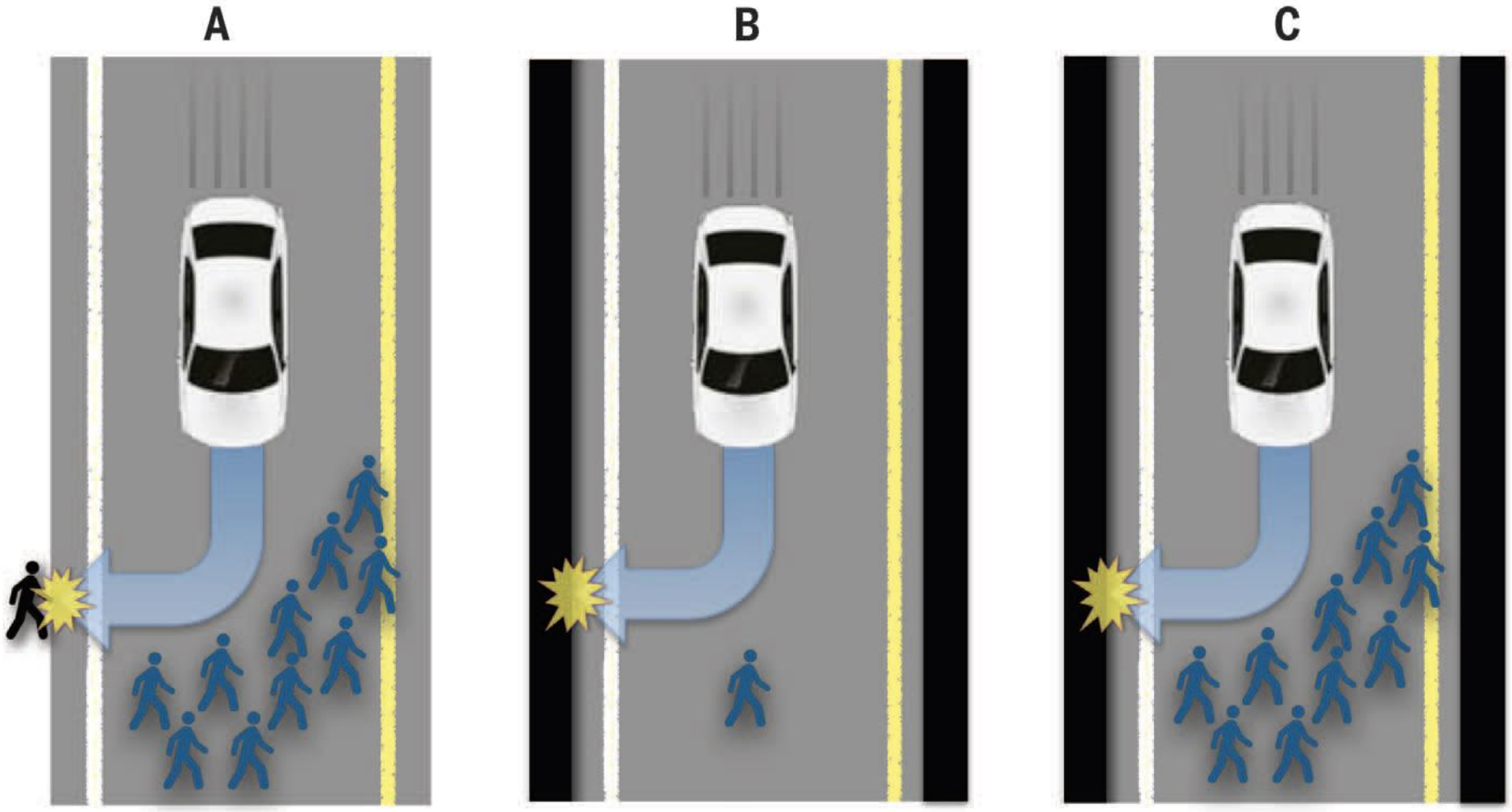


the loop
Cerin, 1987



the man in the yard
Unger, 1992

The social dilemma of autonomous vehicles



Bonnefon et al. 2016

Judith Jarvis Thomson: The surgeon case

- A brilliant transplant surgeon has five patients, each in need of a different organ, each of whom will die without that organ. Unfortunately, no organs are available to perform any of these five transplant operations.
- A healthy young traveler, just passing through the city in which the doctor works, comes in for a routine checkup. In the course of doing the checkup, the doctor discovers that his organs are compatible with all five of his dying patients.
- Suppose further that if the young man were to disappear, no one would suspect the doctor. Do you support the morality of the doctor to kill that tourist and provide his healthy organs to those five dying people and save their lives?

MAI4CAREU | Master programmes in Artificial Intelligence 4 Careers in Europe

Thanks for your attention!

giovanni.sartor@unibo.it

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423