Module 2: Utilitarian Ethics and self-driving cars

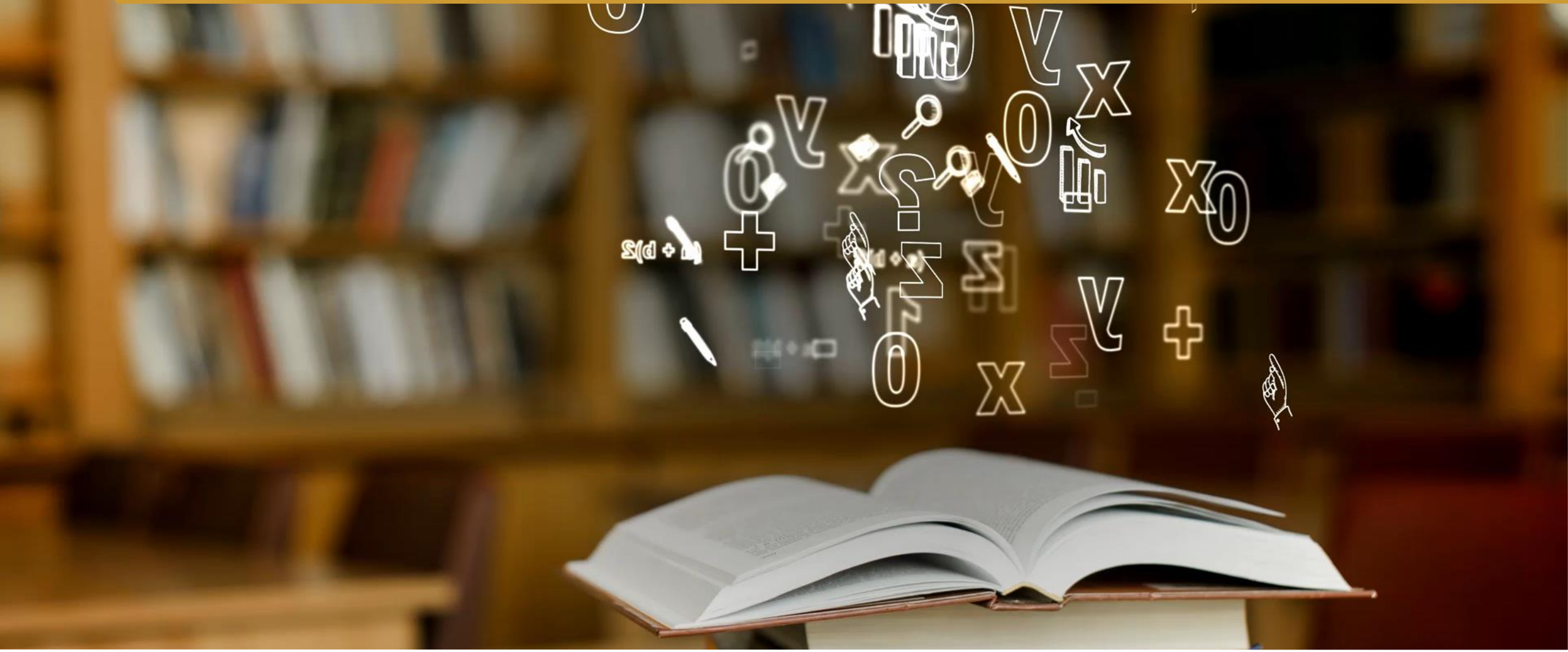
Dr. Gert Meyers (TILT, Tilburg University)

Today

- Utilitarianism
- Blogpost assessment
- The trolley problem and self-driving cars

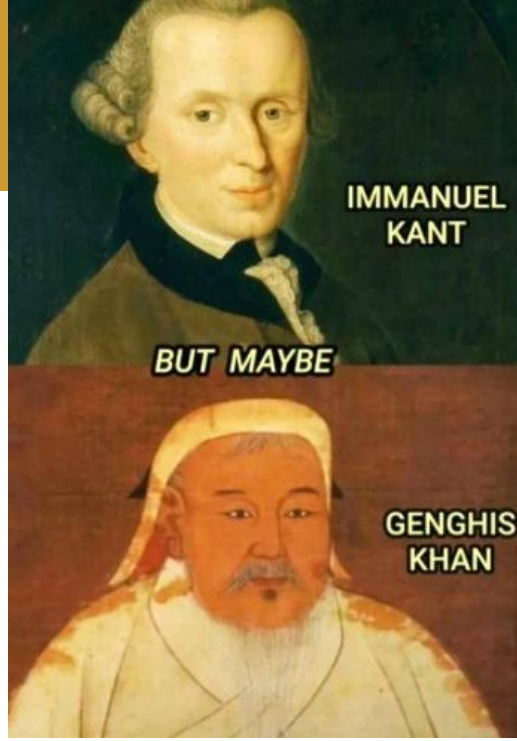
- Does everyone find a debating group?
- Registration debating skills workshop
- Change of plans:
  - 8/4 Autonomy and the ethics of behavioural change
  - 15/4 Trust and trustworthiness in a data-driven context

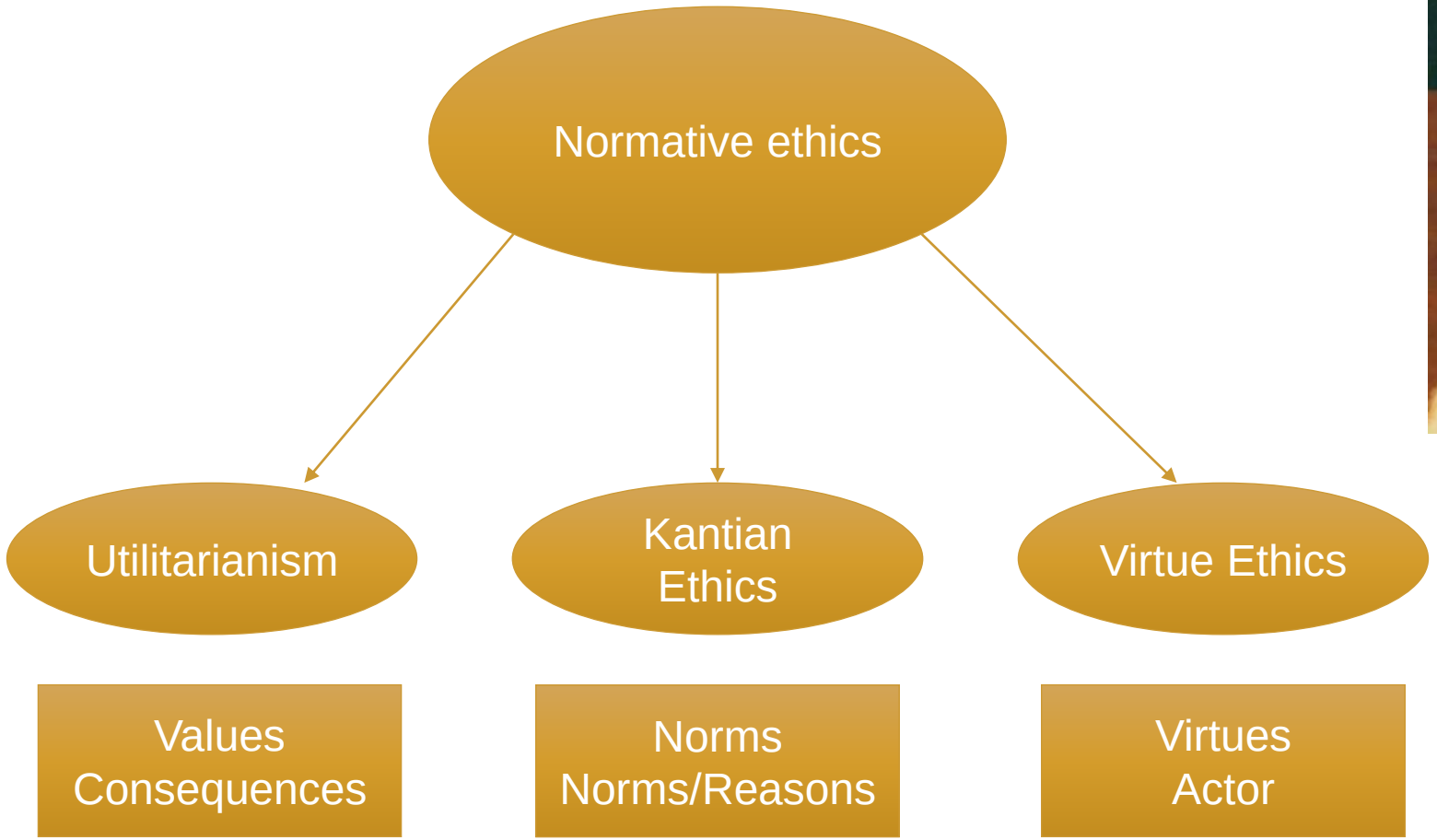
Questions on the course/literature/module 1?



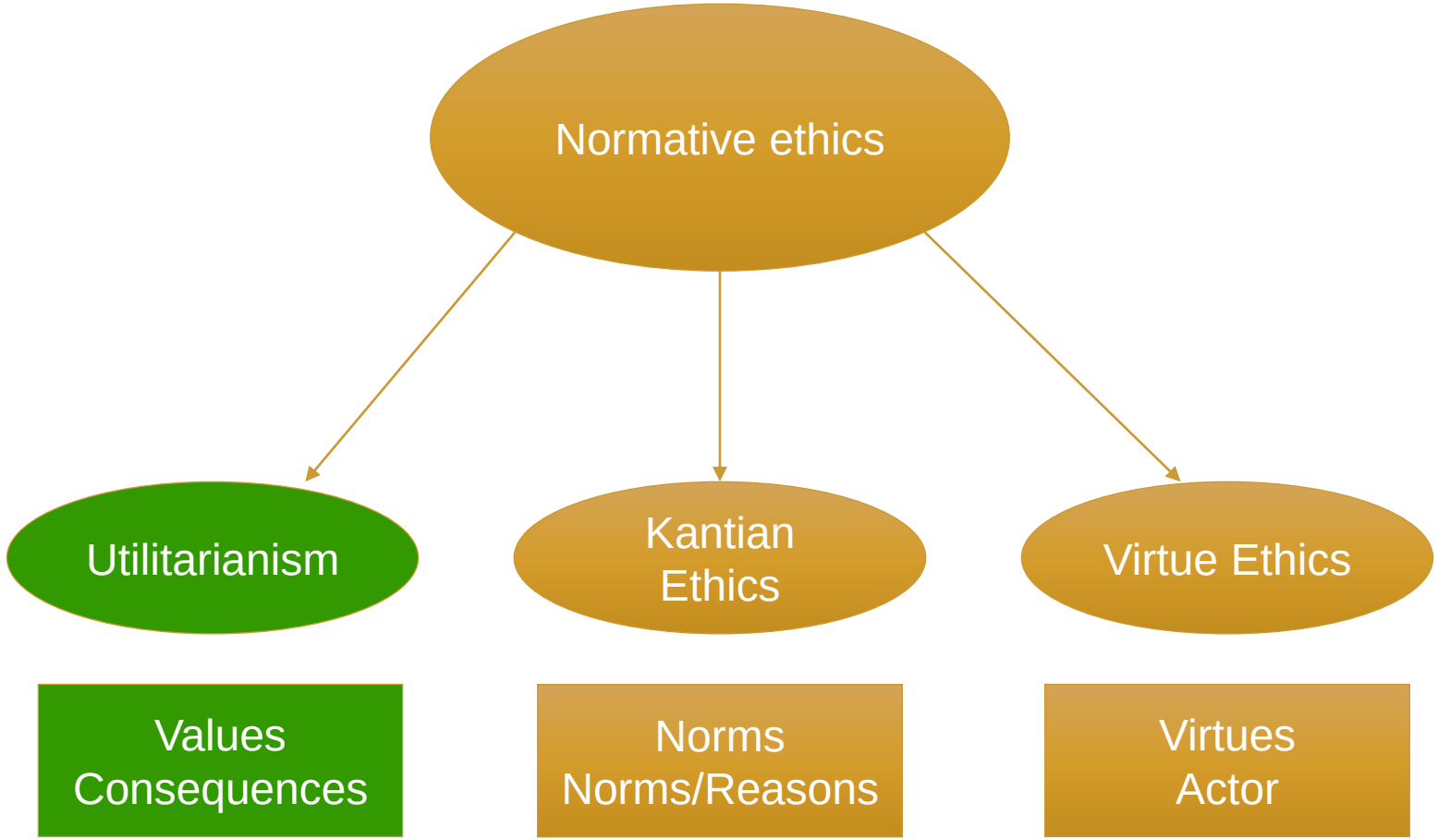
Tilburg University

Ethical theories



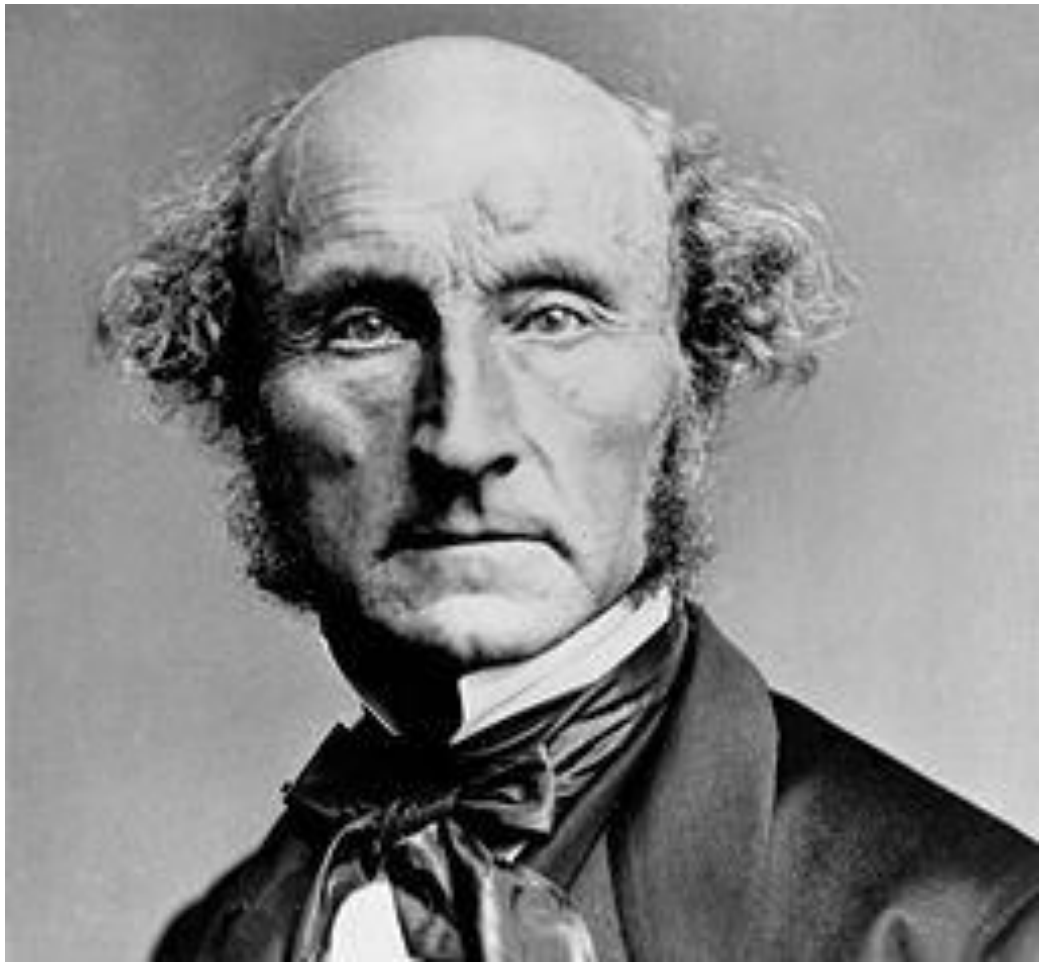


Ethical theories



Utilitarianism: Founding fathers

Jeremy Bentham (1748-1832)



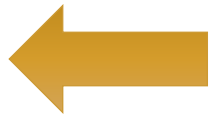
John Stuart Mill (1806-1873)

Utilitarianism (2)

- **Utilitarianism** is a form of **consequentialism**
- **Consequentialism**: the class of ethical theories which hold that the consequences of actions are central to the moral judgment of those action
- **Utilitarianism**: the morally right action is the action that produces the most ‘good,’ where good is understood in terms of ‘**utility**’ or ‘well-being’.
  - Cf. Hedonism (Bentham)

Utilitarianism (3)

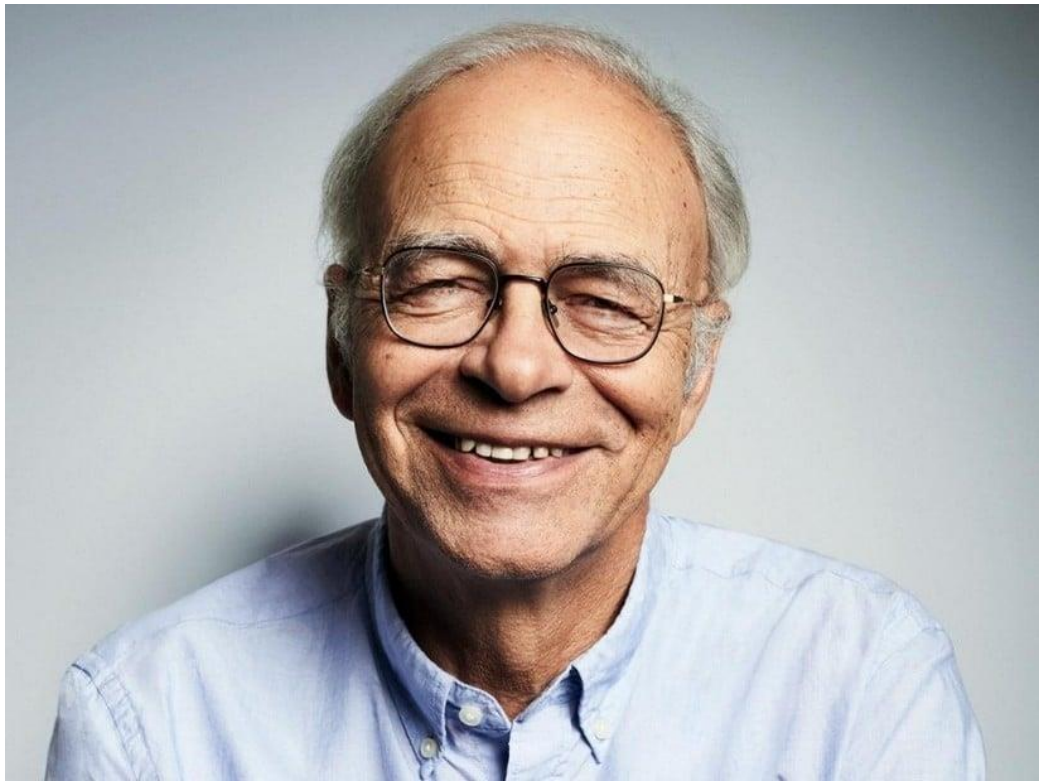
- **Utility principle:** the greatest happiness for the greatest number (of the members of the community).
- Utilitarianism provided/provides the basis for progressive political projects;
  - Bentham: animal rights
  - Mill: women rights
  - Singer: animal rights & duties towards the global poor (impartiality)



- Each person's well-being counts **the same**.
- Right actions are those that produce the greatest possible balance of happiness over unhappiness (e.g.), with each person's happiness counted as equally important.

Utilitarianism (4): animals

- example: Eating of meat
- Meat production causes great suffering for the animals
- We have other options
- The good that is done does not outweigh the evil
  - Interests of non-human animals count!
  - Focus lies on the suffering



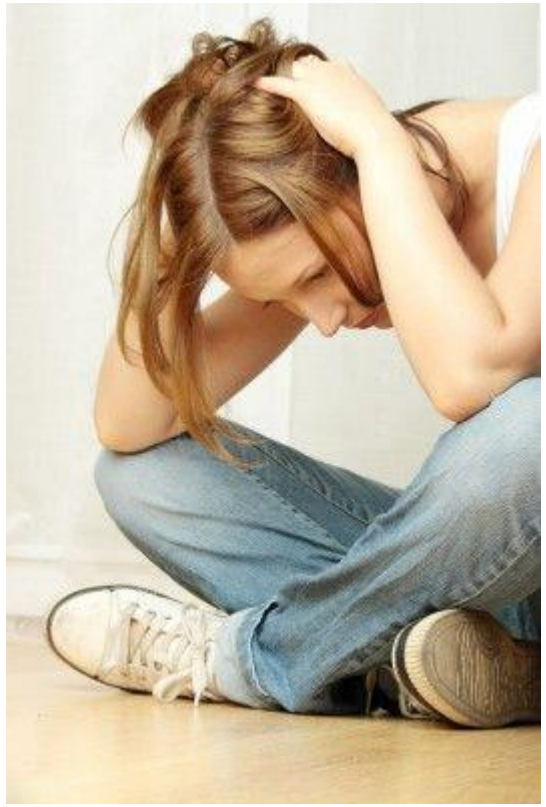
Peter Singer

Utilitarianism (5)

- How to calculate the greatest good for the greatest number?
  - Pleasure is a subjective term
  - How to compare actions
- Freedom principle
  - Everyone is free to strive for their own pleasure, as long as they do not deny or hinder the pleasure of others
- No harm principle
  - Everyone is free to do what one wishes, as long as no harm is done to others

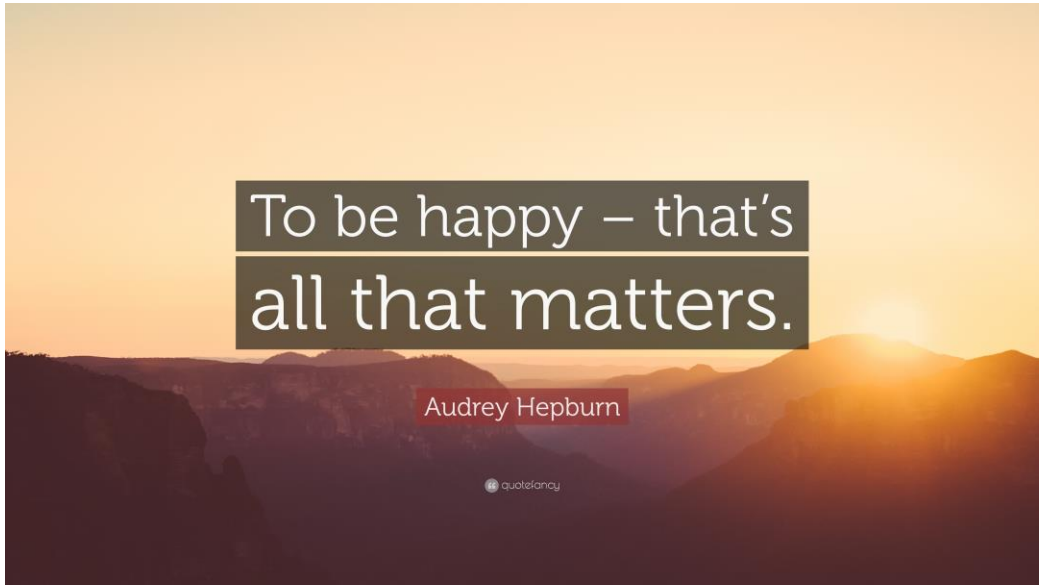
Utilitarianism (6)

- **Act** Utilitarianism: an action is morally right if this specific action maximises utility.
  - Actor calculates for each and every action whether or not it maximises utility
- **Rule** Utilitarianism: an action is morally right if this action is in accordance with a rule whose general acceptance leads to maximising utility.
  - Actor indirectly applies the principle of utility to a specific action by first specifying the norms/rules.



Utilitarianism (7)

- Should we equally be concerned for everyone?
  - Too demanding?
  - What about our loved ones?
- Are consequences all that matter?
  - Justice and societal inequalities
  - Personal rights



Individual assignment: blogpost

- Identify an ethically relevant topic in a **news article** relating to data practices.
- Shortly explain what is morally relevant in this news item
- Reflect on this topic by **employing one of the three ethical theories**.
- Practice in sharing your ethical reflection and conclusions with a wider audience.

- Deadline: March 25th 9.00 A.M., submission via Canvas
- 30% of the total grade
- 500-750 words

“This is why everyone hates moral philosophy professors”

Tilburg University

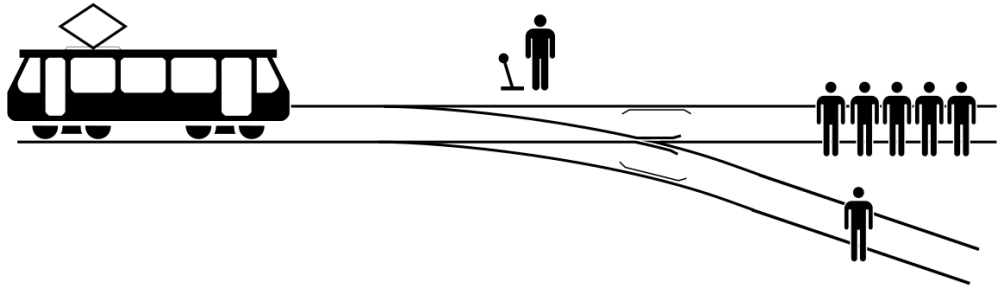
The trolley problem

- There is a runaway trolley down the railway tracks, where there are five people, unable to move. The only way to save five people on some traintracks is to redirect a train to a side-track where, unfortunately, one other person is located. You have two options:

  1. Do nothing, and the trolley kills the five people on the main track.

  2. Pull the lever, diverting the trolley onto the side track where it will kill one person.

- What should you do?
- What would the Utilitarian do?



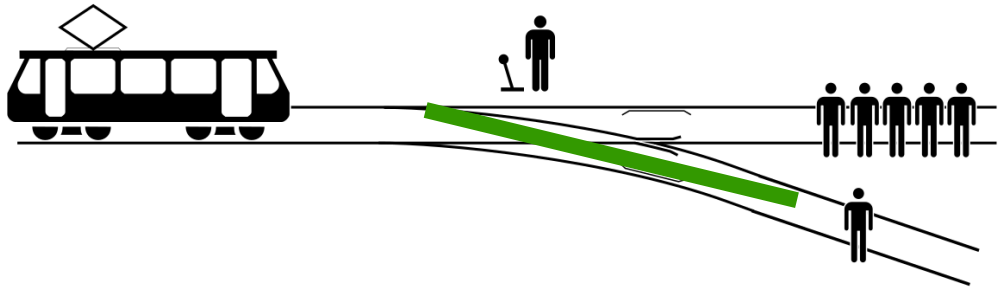
Philippa Foot

Judith Jarvis Thomson

TILBURG UNIVERSITY

The trolley problem (2)

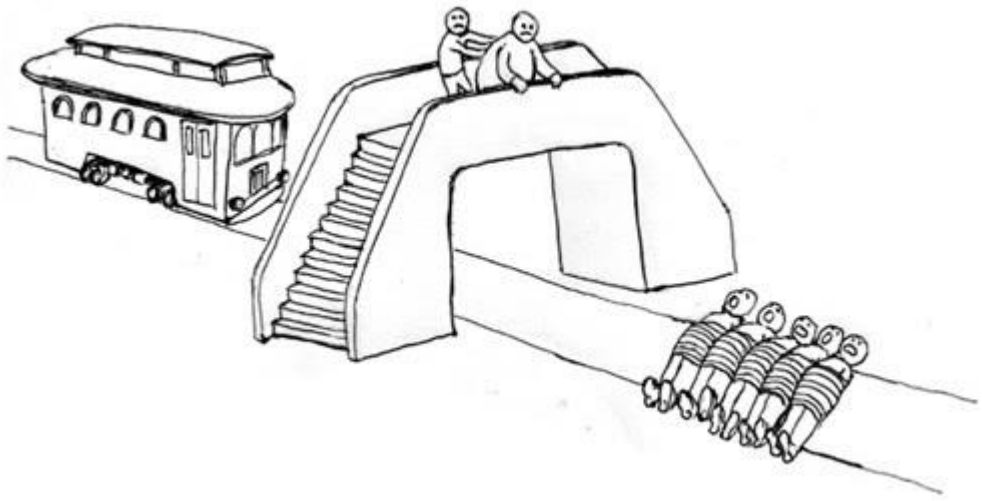
- The ‘Utilitarian’ choice: it’s obligatory to steer to the track with one man on it. According to classical utilitarianism, such a decision would be not only permissible, but, morally speaking, the better option.




'BITS OF THEM FLEW INTO MY MOUTH'
The Good Place

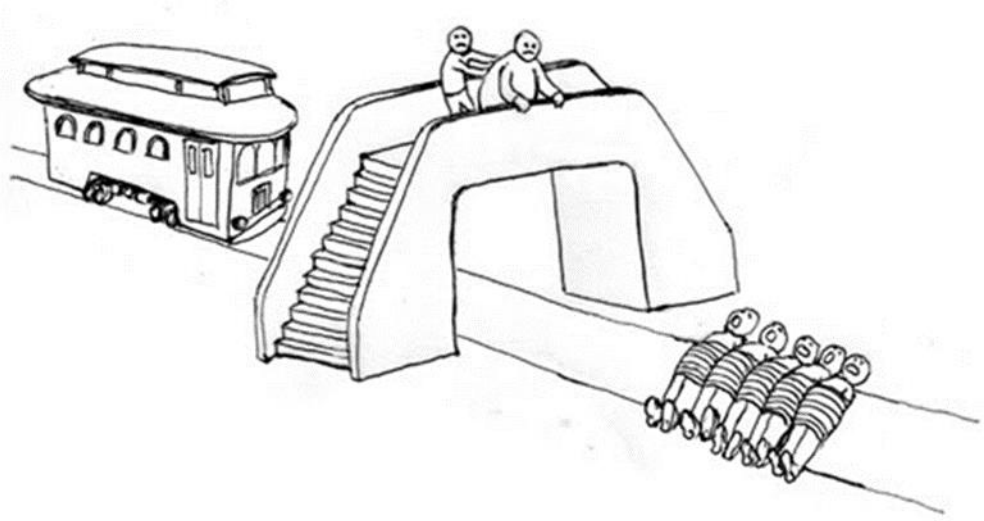
The trolley problem: a variation

- The obese man (Judith Jarvis Thomson)
- ‘As before, a trolley is hurtling down a track towards five people. You are on a bridge under which it will pass, and you can stop it by putting something very heavy in front of it. As it happens, there is a very fat man next to you – your only way to stop the trolley is to push him over the bridge and onto the track, killing him to save five. Should you proceed?’

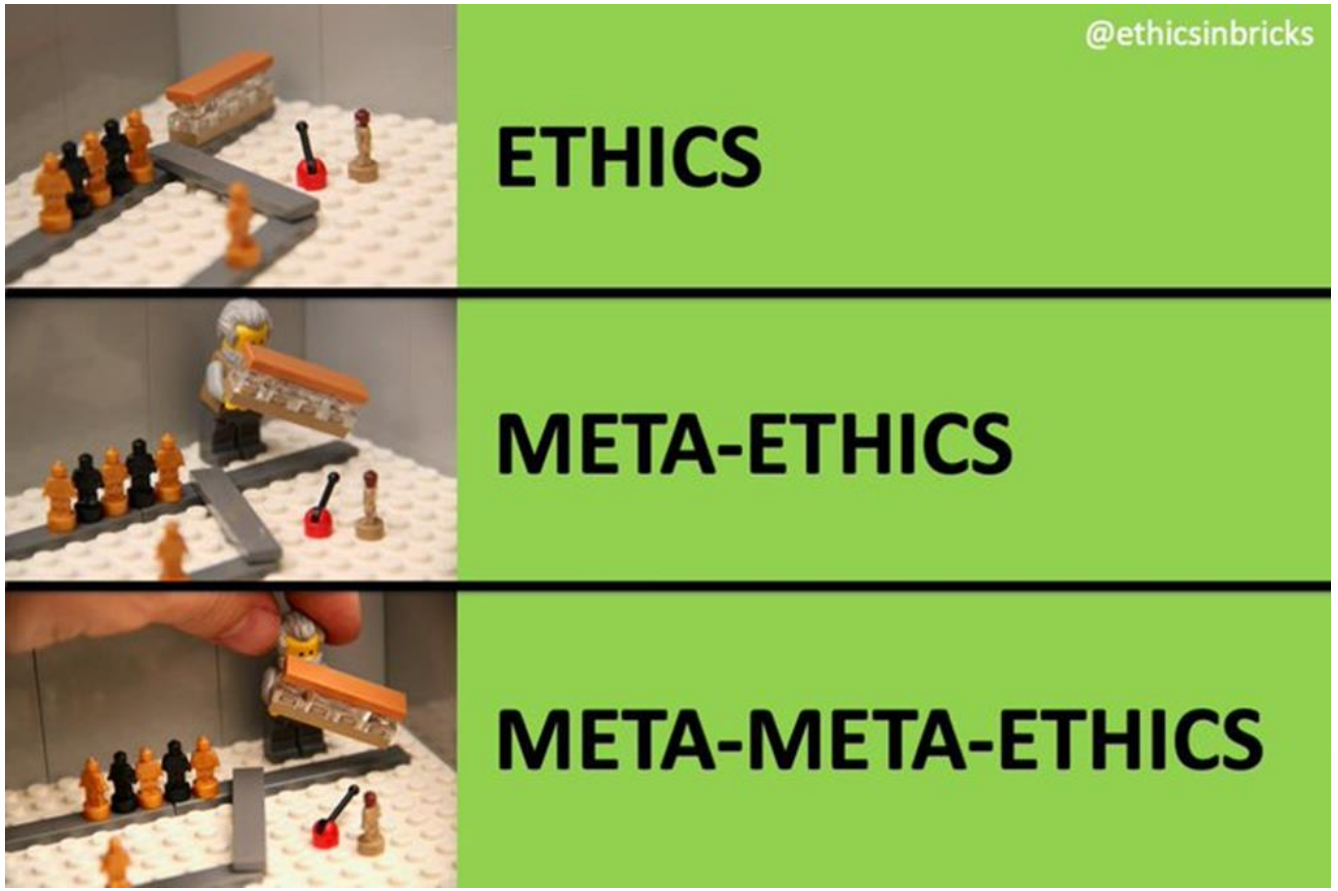


The trolley problem: a variation

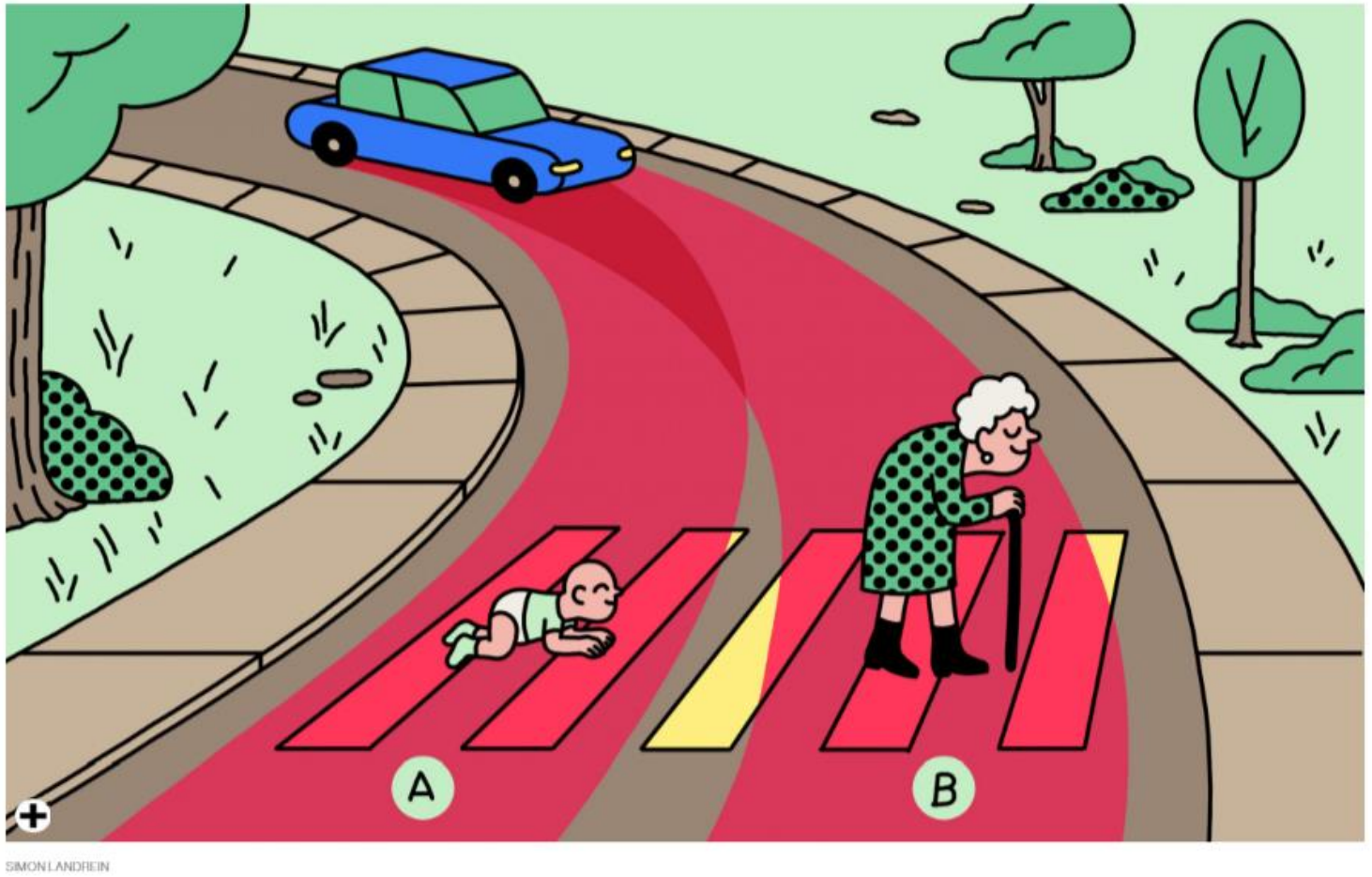
- So: in the ‘classic’ case, many people have the intuition that one should kill the one person to save the many, but in the obese variation, many people have different intuitions. Two questions:
  - We seem to have *inconsistent* moral intuitions. How can we justify why one action (pulling the switch) that would lead to someone’s death is justified while another (throwing a man off a bridge) is not?
  - Do Utilitarians have the only/better cards to make sense of trolley cases, or of securing our intuitions?



Endless trolley problem variations



Self-driving cars and the Trolley Experiment



[The moral machine experiment: video]

Self-driving cars and the Trolley Experiment (2)

Self-driving cars

- Real life consequences
- Pre-programmed by groups/multiple stakeholders
- Moral and legal responsibilities
- Risks, probabilities and uncertainties

Trolley problem

- Thought experiment
- Split second decision by one individual
- ~~[Moral and legal responsibilities]~~
- Stipulated facts and certainties

	Accident-algorithms for self-driving cars:	Trolley Problem:
1a: Decision faced by:	*Groups of individuals/ multiple stakeholders*	*One single individual*
1b: Time-perspective:	*Prospective decision/ contingency planning*	*Immediate/"here and now"*
1c: Numbers of considerations/ situational features that may be taken into account:	*Unlimited; unrestricted*	*Restricted to a small number of considerations; everything else bracketed*
2: Responsibility, moral and legal:	*Both need to be taken into account*	*Both set aside; not taken into account*
3: Modality of knowledge, or epistemic situation:	*A mix of risk-estimation and decision-making under uncertainty*	*Facts are stipulated to be both certain and known*

Next week no clas (next class 19/2): Kantian Ethics and algorithmic accountability

- Ibo van der Poel & Lamber Royakkers (2011). Kantian Theory. /courses/15045/files/2858300?wrap=1
- Reuben Binns (2018). Algorithmic Accountability and Public Reason. https://link.springer.com/article/10.1007/s13347-017-0263-5Links to an external site.

Questions on Module 2?

G.Meyers@tilburguniversity.edu