

Консеквенциализъм

Giovanni Sartor



Концепцията за консеквенциализма

- Действието е морално необходимо
 - ако осигурява най-добрия резултат в сравнение с неговата алтернатива
 - ако добрите му резултати надвишават негативните му резултати в най-голяма степен
 - ако произвежда най-висока полезност?
- Моралът като оптимизационен проблем!
- Различни видове консеквенциализъм
 - Кои са добрите и лошите неща, които трябва да се максимизират;
 - Колко са;
 - Колко всеки от тях има значение;
 - Можем ли да конструираме една функция на полезност, която комбинира печалби и загуби за множество ценни цели;

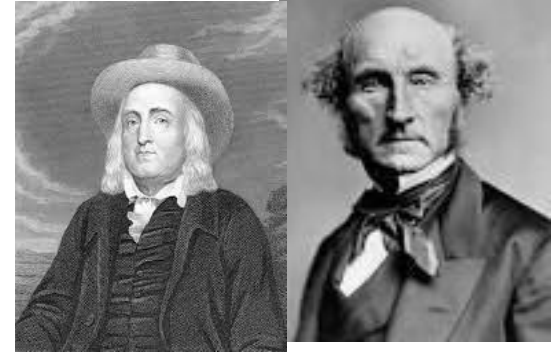
Референтният подход: Утилитаризъм

- Jeremy Bentham,
- John Stuart Mill. От Utilitarianism (1861). Принцип на полезност:
 - Действията са правилни в пропорция, тъй като са склонни да насърчават щастието, погрешни, тъй като са склонни да произвеждат обратното на щастието. Под щастие се разбира удоволствието и липсата на болка; под нещастие, болка и лишаване от удоволствие
- Полезност: Щастие или задоволяване на желаниа/интереси
- Утилитаризмът не е егоизъм
 - Полезността на всички трябва да се взема предвид еднакво



Предимства на утилитаризма

- Концептуално просто
- Егалитарен (полезността на всеки се брои по един и същи начин)
- Съвпада с някои основни интуиции (добре е да правиш хората щастливи, лошо е да ги караш да страдат)
- В много случаи е работещо, в някои случаи проблематично (какво да правим с глада, как да се отнасяме към приятели и роднини и т.н.)



Две версии на утилитаризма

- Действащ утилитаризъм
 - Извършете действието, което увеличава максимално полезността
 - Направете оптимичното действие
- Управляващ утилитаризъм
 - Следвайте правилото, чието последователно прилагане увеличава максимално полезността
 - Следвайте оптимичното правило
- Утилитарен ли е AI
 - Коя функция на полезност би била утилитарна?
 - Трябва ли AI системите да приемат утилитарна функция за възнаграждение?
 - Трябва да следват версиите на Закона или на Правилата (архангели ли са или проли)?

Проблеми с действащия утилитаризъм

- Осигурява ли добра процедура за вземане на решения
 - Можем ли да изберем какво да правим, като оптимизираме резултата от нашите действия? Имаме ли информацията, за да направим това изчисление? Може ли една AI система да има тази информацията?
- Осигурява ли добър стандарт за оценка на решенията?
- Каква е връзката между полезността и функцията за възнаграждение?

Действащ утилитаризъм: Проблеми

- Твърде взискателен ли е;
 - Трябва ли да дам на бедните всичко над минимума, който ми позволява да оцелея?
 - Трябва ли да отдавам еднакво значение на всички, независимо от връзката им с мен?
 - Добре ли е да се навреди на някои хора за по-голяма полза на други
 - Репресии? Изтезание? Садизъм?
- Какво може да каже един утилитарист:
 - Случаите, в които утилитаризмът изглежда се проваля, не са реалистични
 - Няма истински контраст между утилитаризма и основните морални вярвания

Управляващ утилитаризъм

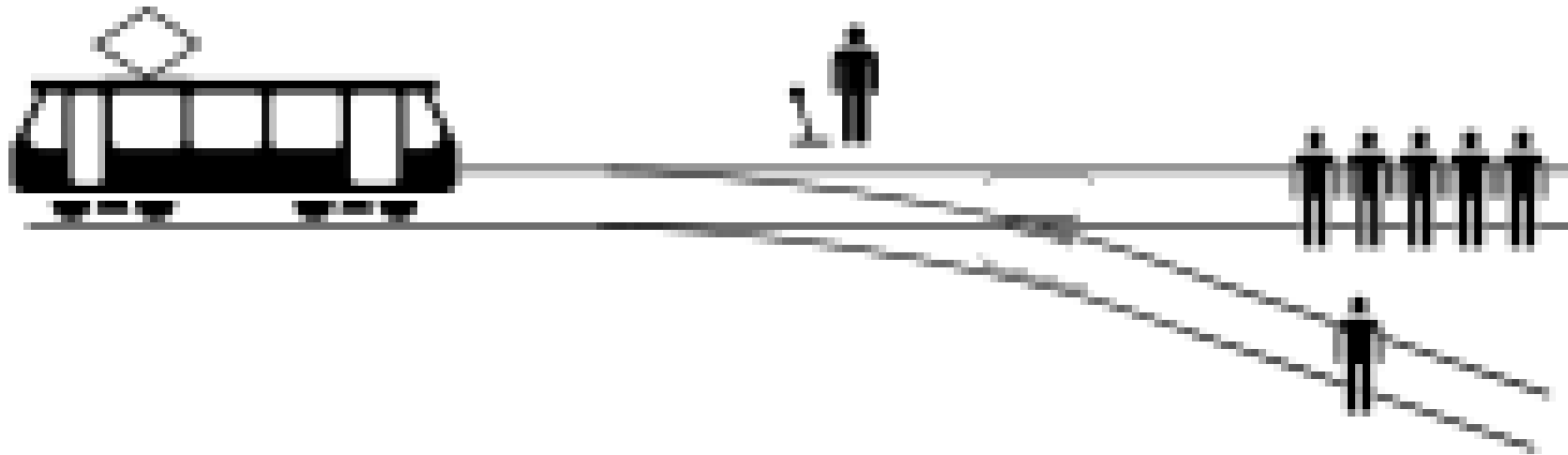
- едно действие е морално правилно само защото се изисква от оптимистично социално правило (социално правило, общото спазване на което би осигурило най-висока полезност)
 - Добре е да се казва истината, да не се краде и т.н., тъй като общото спазване на такива норми би донесло най-голяма полза
 - А какво остава за онези изключителни случаи, в които правилото не дава резултати
 - Какво знаете, че повечето хора не спазват правилото.
 - Трябва ли да бъдем честни, ако повечето хора наоколо са нечестни?

Друг проблем: разпределението

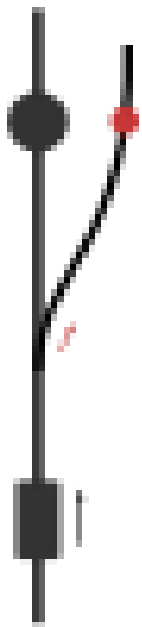


- Има ли значение как се разпределят добрите и лошите резултати?
 - Добре ли е да се направи действие, което е от полза за едни в ущърб на други?
 - Винаги, ако ползите са повече от недостатъците?
- Утилитаризъм срещу максимизиране на богатството
 - Утилитаризмът благоприятства (скромното) преразпределение на богатството, тъй като една и съща сума пари дава повече полезност на бедните, отколкото на богатите
 - Трябва обаче да се вземе предвид въздействието на преразпределението върху генерирането на богатство
- Максимизирането на богатството (възприето от някакъв икономически подход) има за цел да увеличи максимално богатството в обществото, независимо от разпределението.

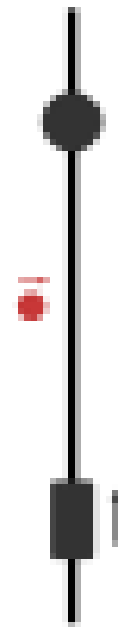
Проблемът с тролея



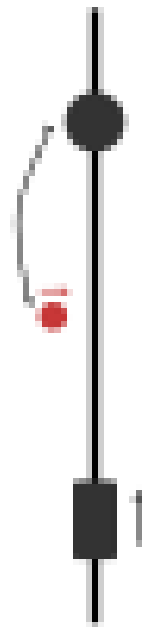
Какво би направил? Какво трябва да прави една AI система, натоварена с задачата да следи трафика?



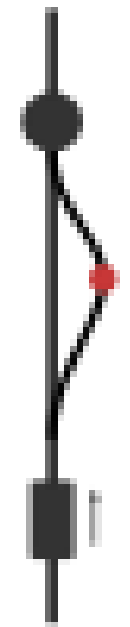
the switch
Farr, 1847



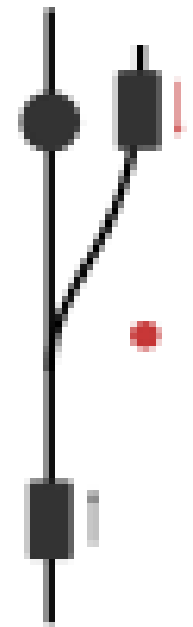
the fat man
Sturmann, 1876



the devil

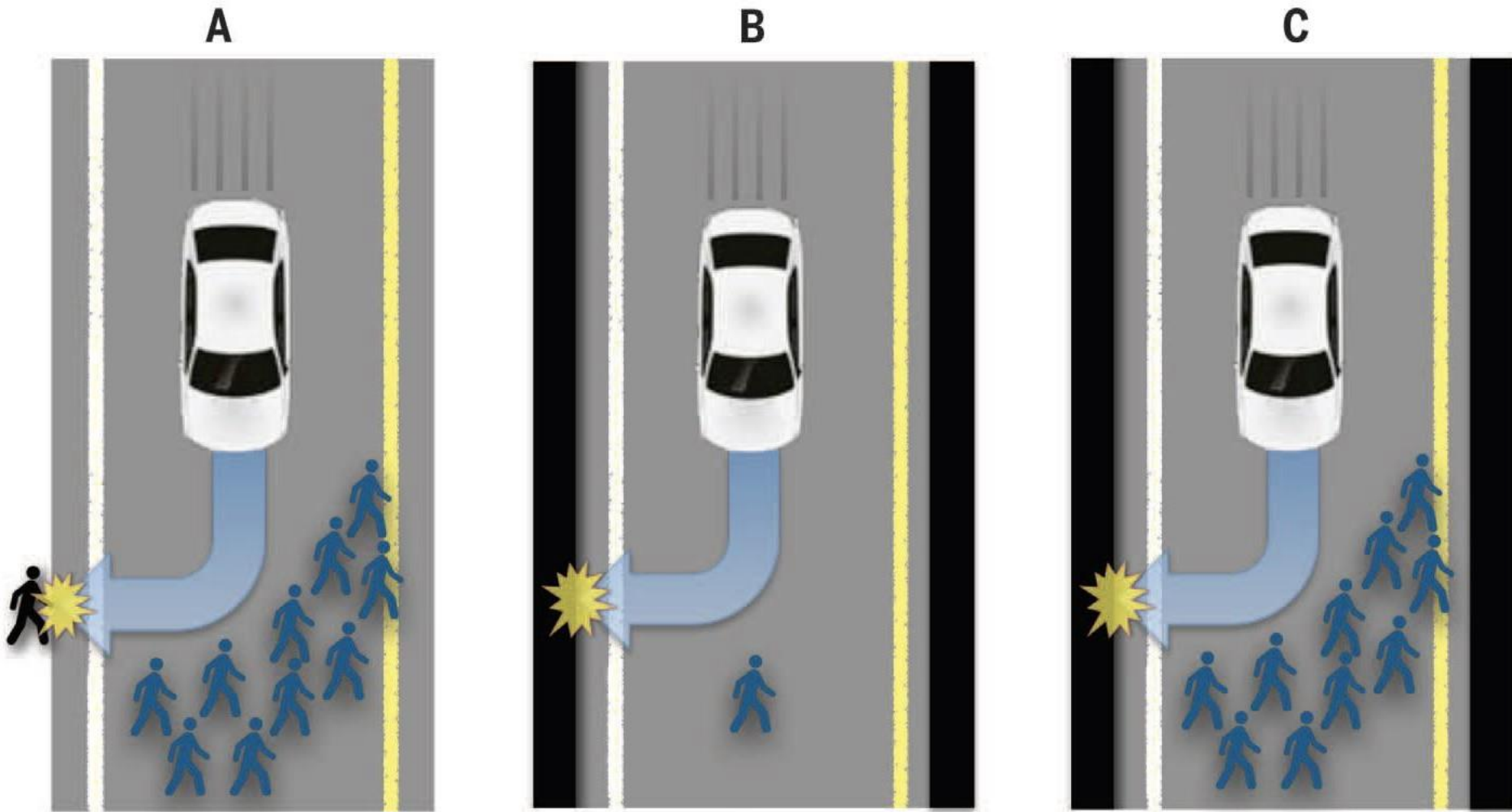


the loop
Cassidy, 1907



the man in the yard
Koger, 1992

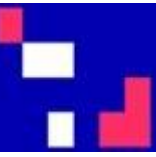
Социалната дилема на автономните превозни средства



Judith Jarvis Thomson: Случаят с хирурга



- Брилянтен хирург по трансплантация има петима пациенти, всеки от които се нуждае от различен орган, всеки от които ще умре без този орган. За съжаление няма налични органи за извършване на нито една от тези пет трансплантационни операции.
- Здрав млад пътешественик, който просто минава през града, в който работи лекарят, идва за рутинен преглед. В хода на прегледа лекарят открива, че органите му са съвместими с всичките му петима умиращи пациенти.
- Да предположим още, че ако младият мъж изчезне, никой няма да заподозре доктора. Подкрепяте ли морала на лекаря да убие този турист и да предостави здравите му органи на петимата умиращи и да спаси живота им?



Благодаря ви за вниманието!

giovanni.sartor@unibo.it

