

Utilitarianism & Social Contract Theory

Lecture 2-3

Computers & Society (CPSC 430)

Kevin Leyton-Brown

<https://www.cs.ubc.ca/~kevinlb/teaching/cs430>

Recap

- What is the categorical imperative?
- What is a utility function?
- How can you justify choices of utility numbers?
- How can you make a utilitarian argument when you don't know exact utility numbers?
- How do utilitarian arguments handle situations when actions can have uncertain consequences?

Act Utilitarianism

- Act utilitarianism
 - Add up change in happiness of all affected beings following from a given action
 - $\text{Sum} > 0$, action is good
 - $\text{Sum} < 0$, action is bad
- How useful is act utilitarianism?
 - discuss with the person beside you
 - report back

Case for Act Utilitarianism

- Focuses on happiness, which is intuitive
- Down-to-earth (practical)
 - Straightforward to apply
 - Can therefore be helpful in resolving disputes
 - decomposes big issues into lots of little questions
- Comprehensive
 - Allows an agent to trade off different aspects of a situation
 - Contrast with Kantianism: we needed to find one rule

Case Against Act Utilitarianism

- Unclear whom to include in calculations
 - Utilitarians might say you should never exclude anyone...
- Too much work
 - But it's OK to follow a “rule of thumb” most of the time.
- Ignores our innate sense of duty
 - Suppose I make a promise, but can get \$1 for violating it.
 - Seems to miss the sense that I care about my word.
 - Quinn claims: “Note that it does no good for an act utilitarian to ... say that the hard feelings caused by breaking my word to A will have a negative impact on total happiness of $-N$ units, because then all I have to do is change the scenario so that breaking my promise to A enables me to produce $1,001 + N$ units of good for B. We've arrived at the same result.”
 - But is this a problem?
- Susceptible to the problem of moral luck
 - Whether an action is moral depends on outcome, which can depend on circumstances beyond your control

Rule Utilitarianism

- We ought to adopt **moral rules** which, if followed by everyone, will lead to the greatest increase in total happiness
 - Act utilitarianism applies Principle of Utility to individual actions
 - Rule utilitarianism applies Principle of Utility to moral rules

Case for Rule Utilitarianism

- Not every moral decision requires utilitarian calculus
 - You only have to work out the morality of rules.
- Moral rules survive exceptional situations
 - A rule utilitarian can reason (a bit like a Kantian) that it's better for everyone to keep their promises than to lie, and so reject lying for a \$1 gain
- Avoids the problem of moral luck
 - We look at the overall usefulness of the rule, not individual outcomes

Case Against RU, and Utilitarianism in General

- RU: need to identify a single rule to describe situation
- All consequences must be measured on a single scale
 - All units must be the same in order to do the sum
 - In certain circumstances utilitarians must quantify the value of a human life
 - BUT: good arguments from utility theory
 - We have to figure out what this utility function is!
- Utilitarianism ignores problem of unjust distribution
 - Utilitarianism (as defined here) doesn't mean "the greatest good for the greatest number"
 - That requires a principle of justice
 - We can try to combine these ideas
 - However, what happens when a conflict arises between the Principle of Utility and our principle of justice?

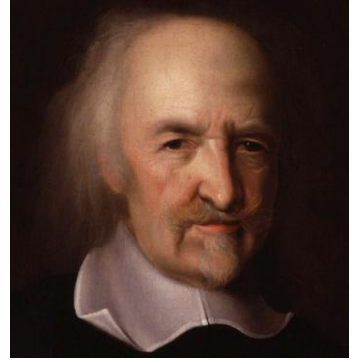
An Example

- Is it ethical to pay a ransomware attacker to unlock my data?
 - Discuss quickly with a partner
- Act Utilitarianism:
 - for me: my data is worth more than the payment so I'm better off paying
 - for the attacker: prefers to be paid
 - for everyone else: me paying the attacker increases the risk of more attacks, but not by much
- Rule Utilitarianism:
 - rule: people should pay ransomware attackers to unlock their data
 - if everyone followed the rule, attacks would be much more common
 - if nobody followed the rule, attacks wouldn't succeed so attackers wouldn't try
 - even though we're individually better off by paying, we're collectively better off if nobody pays

Social Contract Theory

- Thomas Hobbes

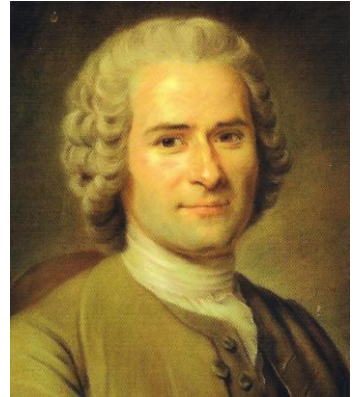
- “State of nature”
- We implicitly accept a social contract
 - Establishment of moral rules to govern relations among citizens
 - Government capable of enforcing these rules



1588-1679

- Jean-Jacques Rousseau

- In ideal society, no one above rules
- That prevents society from enacting bad rules



1712-1778

Social Contract Theory Definition of Morality

James Rachels:

“Morality consists in the set of rules, governing how people are to treat one another, that rational people will agree to accept, for their mutual benefit, on the condition that others follow those rules as well.”

If you're an econ geek:

“every Nash equilibrium in sets of social rules is morally right.”

Kinds of Rights

- **Negative right:**
 - A right that another can guarantee by leaving you alone
- **Positive right:**
 - A right obliging others to do something on your behalf
- **Absolute right:**
 - A right guaranteed without exception
- **Limited right:**
 - A right that may be restricted based on the circumstances
- Positive rights tend to be more limited
- Negative rights tends to be more absolute