

Utilitarianism I

Ethical hacking and hats



Judge others! Avoid excessive tolerance

- Last week we looked at the importance of judging others - in the sense of ***evaluating*** whether actions are right and wrong - in a fair way.
- In this sense, we *should* judge others in our community and in other communities. We should judge friends and strangers.
- Don't passively accept a company's or co-workers' actions and values. Moral judgement is a bit like voting. We don't all get it right, but it's important we all try, and it's our right.
- Appeals to tolerance are often abused, such as:
 - 1. Calling you out for not being tolerant (when you shouldn't be)
 - 2. False moral equivalence (e.g. the other side does the same, so we can do it / everybody does it so it must be OK)

Also avoid hypermoralism

We must also watch out for hyper-moralism. There are moral panics today associated with both the left and right – often translating into defending one good value *at the cost of all others*, transforming value into mistake.

The sometimes obsessive moral policing associated with cancel culture, or the obsession with ending Diversity Equity and Inclusion policies. Defense of freedom as speech as trumping all other rights - freedom to say anything racist, sexist, etc

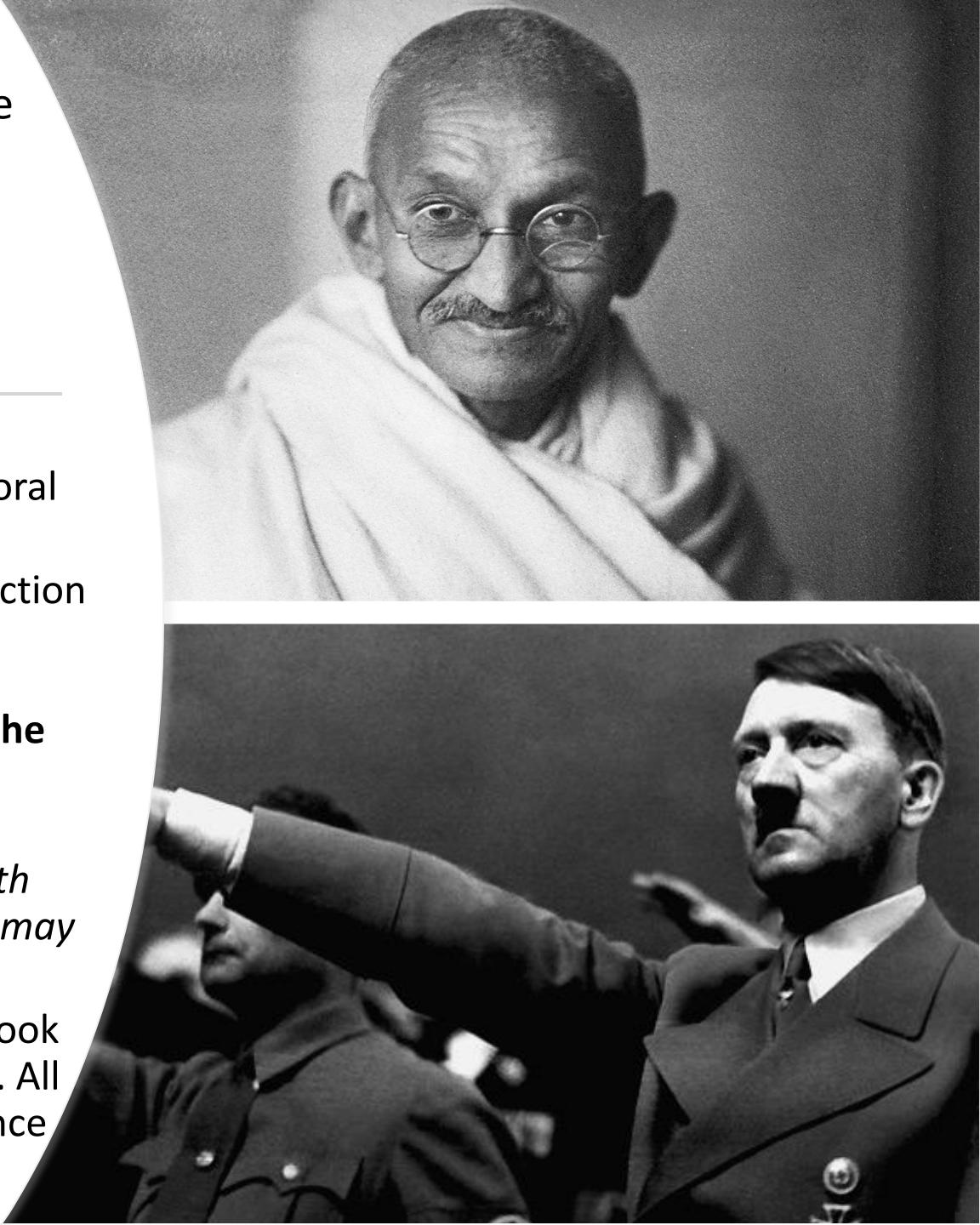
Blanket judgements are typically poor judgements. Make sure you target individual behaviours, rather than rush to conclude a whole person or culture is evil and must be silenced.

To understand you must judge is NOT to throw away the concept of tolerance. It's crucial to be tolerant. As long as we don't tolerate the intolerable. **Tolerance must have limits to be meaningful and useful**



Revisions and moving on to the three main theories of normative ethics

- Q: Someone is an objectivist/relativist. Are their moral values correct?
- A: We don't know. The mere rejection of relativism is not enough to guarantee morality. **See two objectivists on the right side of the screen.**
- If you reject relativism, you now believe there are **some** things with *objective value – but you may or may not know which ones.*
- To know what to do we need to look into normative theories of ethics. All reject relativism + provide guidance
- We'll start with utilitarianism



Revisions

- Normative Relativism: **Culture of agent** defines right/wrong. We can only judge by reference to the culture of the agent (within that relative standard)
- Normative objectivism: Morality does not depend solely on social custom or individual acceptance (i.e. normative relativism is wrong)
- Objectivist & absolutist: Moral objectivism is right + there are **principles blind to circumstances** (e.g. “You ought to keep your promises”, “Hacking is wrong”)
- Objectivist & non-absolutist: Moral objectivism is right + there are objective values (e.g. protect the innocent, honesty, loyalty, **pleasure - utilitarianism**), but they cannot be captured in rules that apply in all cases without regard for **circumstances**
- **Don't confuse a resistance to absolutism (i.e. stating that circumstances are morally relevant) with relativism (i.e. stating that culture/opinion of the observer defines what has value).**

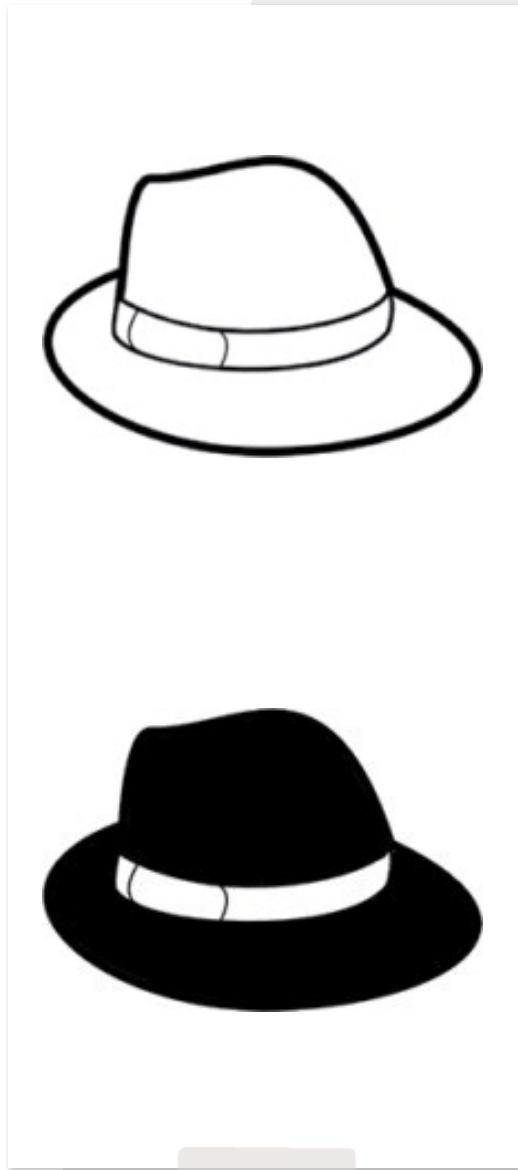


**Culture of agent
(e.g. company values)**

**Circumstances of
moral dilemma (e.g.
does hacking
company X protect
innocents and harm
the guilty?)**

Applied topics

- In the next two weeks we'll focus on Internet fraud as applies examples: cybercrime, fraud, hacking, spoofing, phishing, dark patterns,...
- We'll focus a little more on hacking as an example
- Hacking can be ethically problematic due to a number of reasons
- But not all hacking is the same.



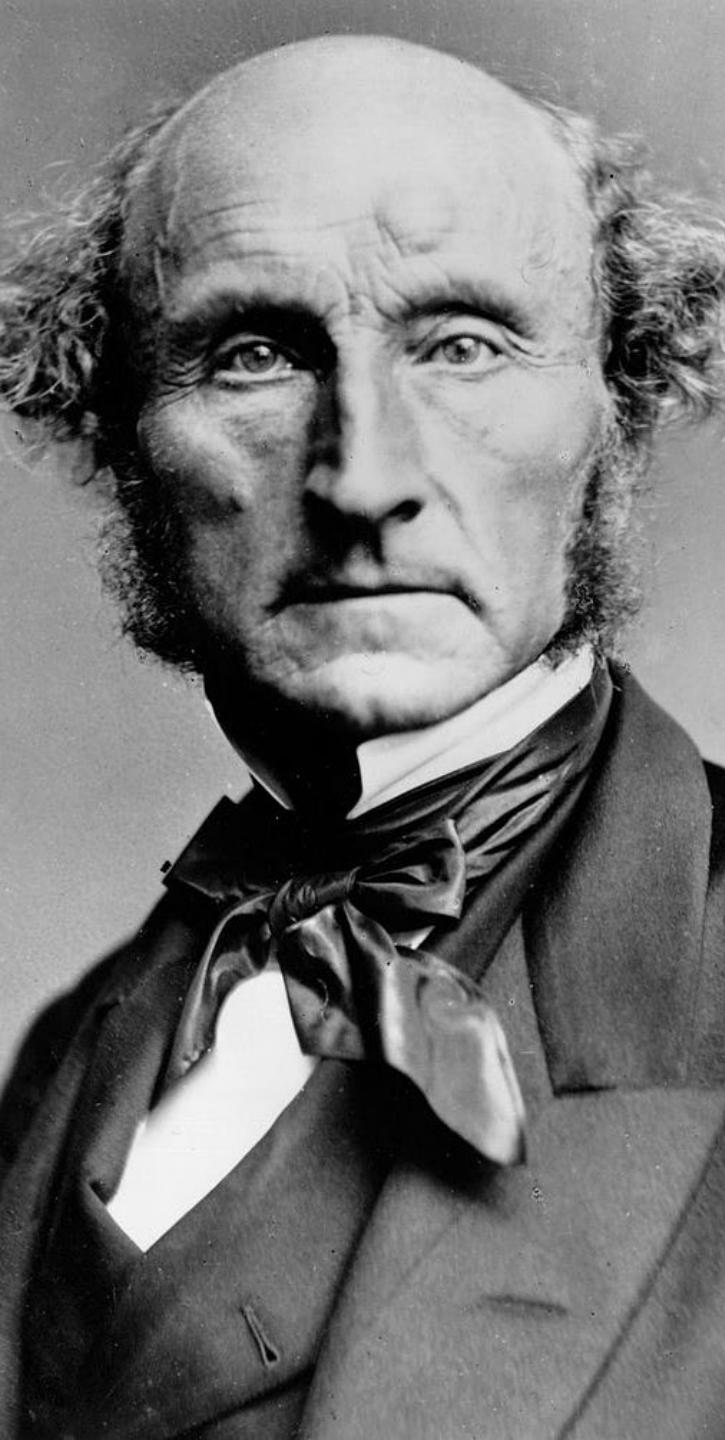


Using utilitarianism to look into hacking

- White hat hacking is done with permission (paid and legal penetration tester)
- Black hat hacking is done without permission and the goal is to steal data, destroy data, make the network unusable, steal confidential info, etc
- Grey hat hackers can break the law, but then not take advantage, or even offer to fix security issue
- **Are any of these ethical?**
- Common sense would tell you:
 - White – Yes
 - Black – No
 - Grey – Possibly – But how can we tell?
 - We'll use utilitarianism to investigate

Utilitarianism and thinking about absolutism

- The main intuition-based distinctions between white, black and grey hat hacking regard:
 - 1. permission/consent (with/without)
 - 2. legality (legal or illegal)
 - 3. intention (to help/harm) and
 - 4. **outcome** (does it actually help/harm?)
- Utilitarianism tells us we should look only at the outcome, so actually any of these types of hacking could be right or wrong, *according to a utilitarian*.
- We'll consider: Is good and bad absolute (e.g. murder, stealing) or can it be weighed against the alternatives? Can you do something usually *seen* as wrong as a morally legitimate (i.e. *justified*) means to an end?
- Examples: grey hat hacking, hacktivism (e.g. WikiLeaks or Anonymous), a experienced black hat converting to a very useful white hat



Summary of the 3 approaches to ethics - chronology

- Aristotle (385-323 BC) and Hume (1711-1776) - **virtue ethics** - Focus on character and moral dispositions (virtues). What dispositions constitute a good life? Actions are moral/virtuous if they are the right actions *and* are performed for the right reason (virtuous motivation).
- Kant (1724-1804) – Kantian ethics, a **deontological** theory (i.e. based on **duty**). The focus is on motivation – duty and rationality. Actions have moral worth if they are the right actions *and* are performed for the right reason (duty).
- **Bentham - (1748-1832)** and **Mill – left - (1806-1873)** – classical act utilitarianism –the focus is on the outcome of an action in terms of pleasure and pain. Locates the moral quality of actions on the *actual effect of actions on others*.

JEREMY BENTHAM



Bentham

- “As well as being a fascinating, radical thinker in the 18th and 19th centuries, he had a walking stick he called Dapple, a teapot he referred to as Dickey and an elderly cat he named The Reverend Sir John Langbourne.”
- But in death, Bentham was even more unusual. He asked that his body be preserved so that he could be wheeled out at parties if his friends missed him.”
- His body is still at UCL, but with a wax head. The head was removed as it was preserved by a friend who wasn’t an expert and didn’t do a great job. The head was also kidnapped by King’s College.
- [Link](#)

Background - Bentham on the complementary roles of laws and morals

- Bentham was a psychological egoist most of his life
- Bentham - For social control/reform, we have two tools: **laws** and **morals** (he thought morals operate based on self-interest, e.g. reputation, reciprocity)
- Where legislation cannot reach or shouldn't reach, we have morals to control behaviour.
- According to Bentham, we can evaluate actions by focusing on the social utility of the **action itself**. **Actions are good if their consequences are useful**
- Because utilitarianism determines morality based on consequences, we call it a **consequentialist** theory of normative ethics.

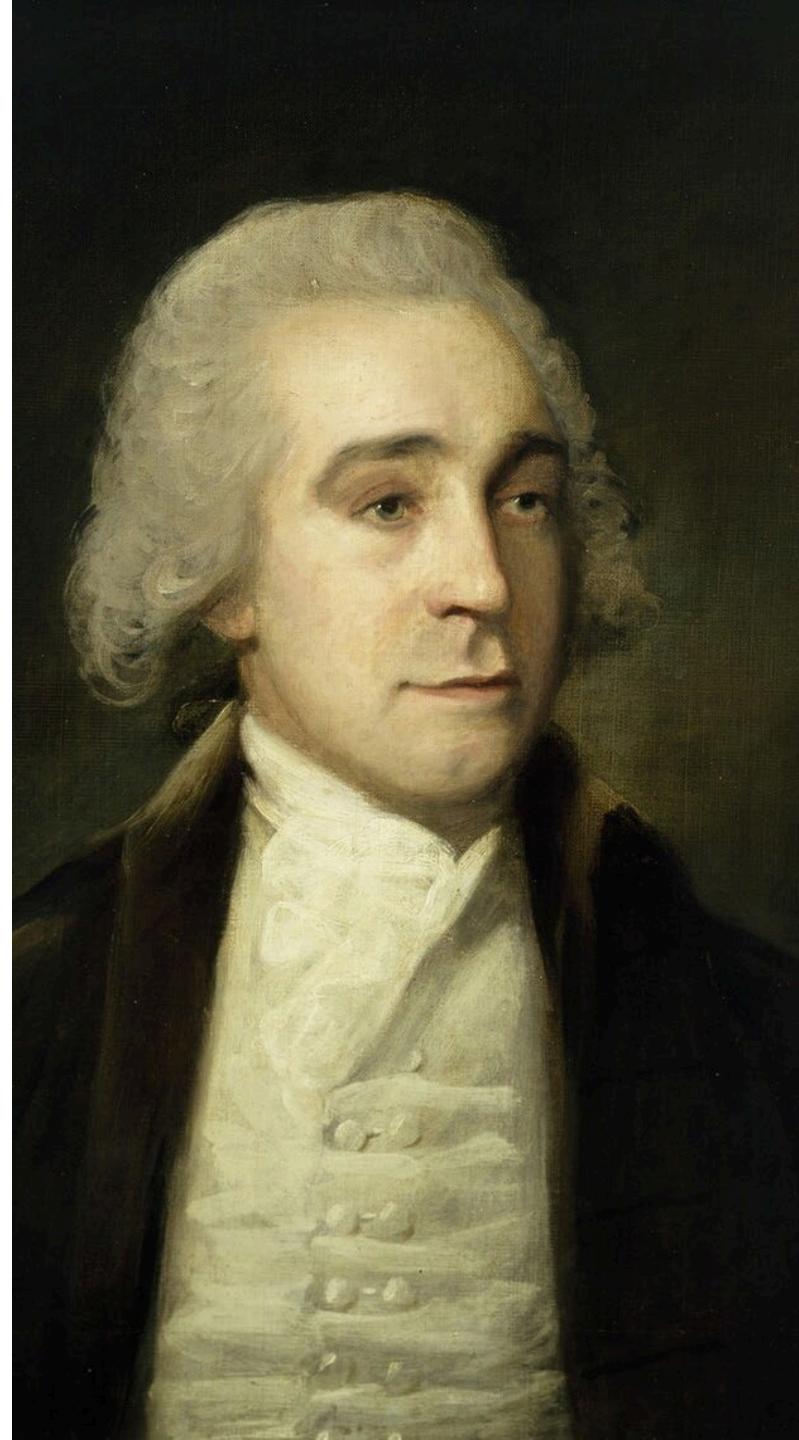
The importance of pleasure in our lives

psychological hedonism and normative hedonism

- Bentham, Introduction to the Principles of Morals and Legislation (1789)
- “Nature has placed mankind under the governance of **two sovereign masters, pain and pleasure**. It is for them alone to point out **what we ought to do**, as well as to determine **what we shall do**. On the one hand **the standard of right and wrong**, on the other the **chain of causes and effects**, are fastened to their throne. **They govern us in all we do, in all we say, in all we think**: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it.” Bentham
- **Psychological/motivational hedonism:** Pleasure and pain are the ultimate *explanations* for action.
- **Normative hedonism:** Pleasure is the only intrinsic good, pain the only intrinsic bad. The tradition comes from Epicurus (Ancient Greece)

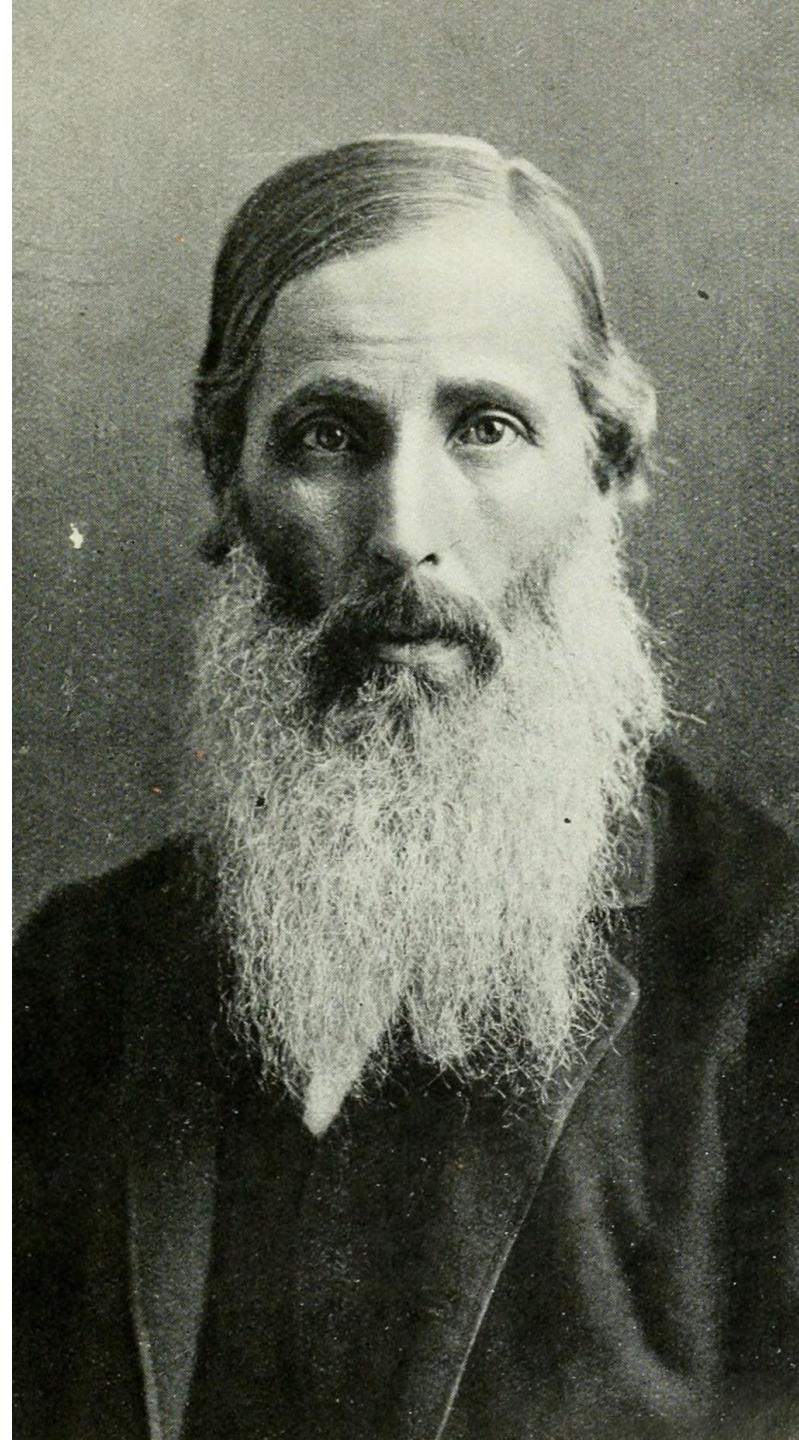
Bentham's principle of utility

- “By the principle of utility is meant that principle which approves or disapproves of every action whatsoever, **according to the tendency** which it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words, to promote or to oppose that happiness.”
Bentham
- Correct interpretation of “according to the tendency”: What we should actually do is what increases happiness/decreases pain **the most** of all available options.



Important utilitarian principle: equal consideration of interests

- **The pleasure or pain of one person should be of no more value than the pleasure or pain of another.**
- Take a perfectly impartial “view from the universe” – Sidgwick
- Not a principle of general equality
- Not a principal of equal treatment or just distribution.
- The action recommended by applying the ECI may well only benefit a few, or only one.
- **The idea of ECI is: The interests of each person affected by the action must count, and they must count equally, when *measuring* utility. Do not consider ourselves and our interests, or those of those close to us as being more important.**



ECI = Interests always counted equally

ECI does NOT mean individuals are always *treated* equally

Utilitarianism is not necessarily about serving the interests of the “majority”



-2

-1

-2

+500

+5

For the sake of this example and for easier visualization, we're purely quantifying pleasure and pain (assigning values). We'll see that Mill rejects Bentham's strict quantification of utility

Clarifying another famous utilitarian principle

Again, not advocating fairness or equal distribution

- Bentham wrote (1843): "Priestley was the first (unless it was Beccaria*) who taught my lips to pronounce this sacred truth:- **That the greatest happiness of the greatest number is the foundation of morals and legislation.**"
- This does **not** mean *just* distribution - that utility should be *evenly divided* among the largest number of people possible. Or that we serve the majority
- **Only total aggregated utility matters.**
- Even though Bentham *personally* thought just distribution is *more likely* to maximize utility, and that we shouldn't sacrifice the rights of one for the greater good - "the interest of everybody is sacred, or the interest of nobody", it's not clear how his *theory* of utilitarianism can satisfy his ideal
- ***Historical note:** It was actually Francis Hutcheson's (1725) who said: "that Action is best, which procures the greatest Happiness for the greatest numbers; and that, worst, which, in like manner, occasions Misery." Hutcheson actually later abandoned this formulation as he considered it flawed.

Quantification of pleasure and pain

(note that this is pre-Mill utilitarianism and was not kept by Mill)

- Bentham's **felicific calculus**, later abandoned by Mill:
- One measures an action's value with respect to pleasure and pain according to the following: **intensity** (how strong the pleasure or pain is), duration (how long it lasts), **certainty** (how likely the pleasure or pain is to be the result of the action), **proximity** (how close the sensation will be to performance of the action), **fecundity** (how likely it is to lead to further pleasures or pains), **purity** (how much intermixture there is with the other sensation). One also considers **extent** — the number of people affected by the action.
- Bentham: we don't have to always *measure*. We often know by experience/common sense that e.g. being in a fight for no good reason will be outweighed by the pain/ Lying causes problems / ...
- Crucially – no type of action (e.g. stealing, lying) is intrinsically right or wrong. This is a rejection of absolutism. Actions can only be wrong when they fail to be appropriate means to maximizing utility (he approved of autonomy and liberty, but *only insofar as they have instrumental value*)

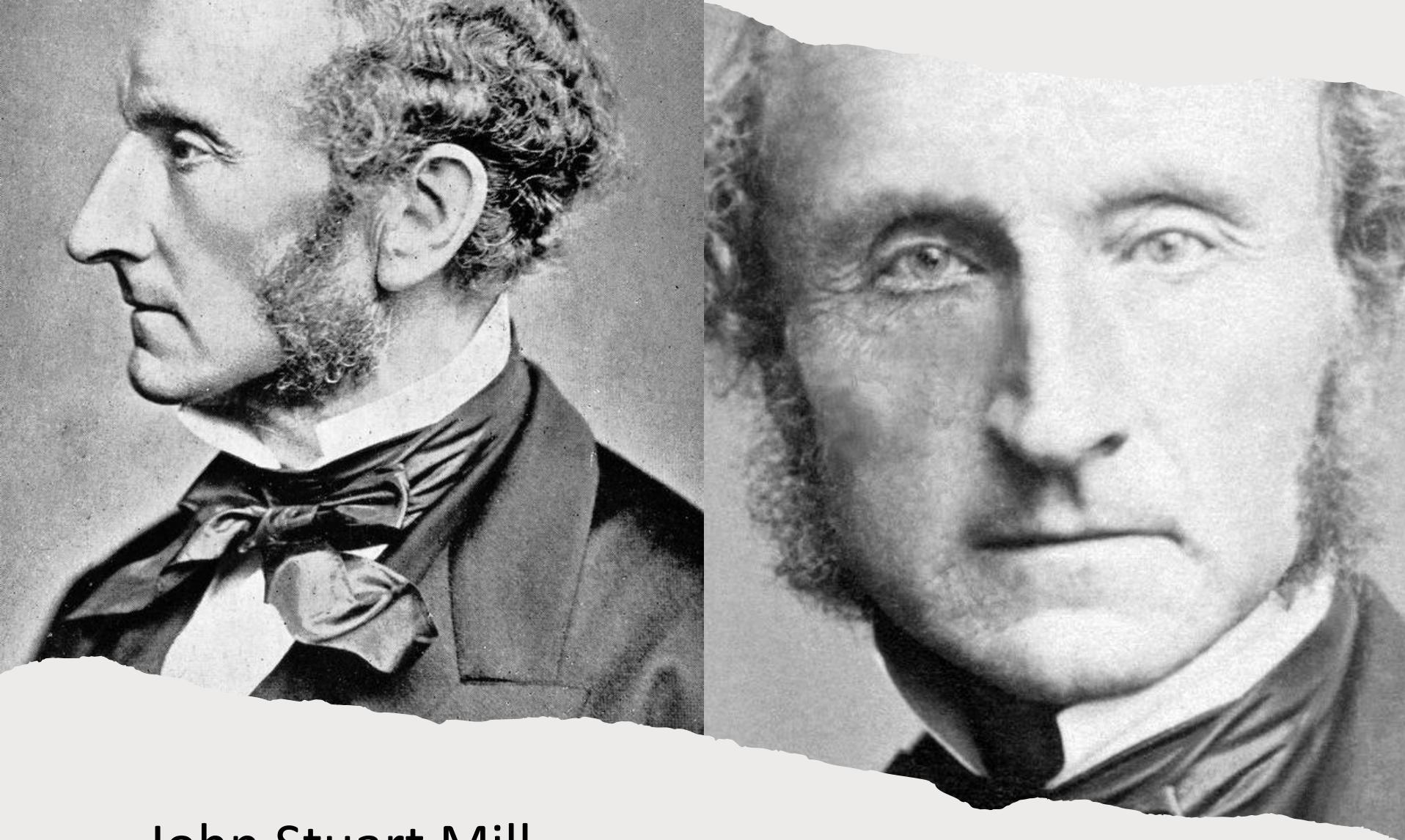
Summary of Bentham's classical act utilitarianism

- 1. **Psychological hedonism.** Highlights the fundamental role of pleasure and pain in human life – the two “sovereign masters” as the only intrinsic motivations
- 2. **Normative Hedonism** – The only intrinsic good is pleasure - Equates the good with pleasure and the bad with pain.
- 3. We should approve or disapprove of an action on the basis of the amount of **utility** (pleasure/pain) it produces as a consequence – it is a **consequentialist** normative theory.
- 4. Pleasure and pain can be **quantified** (felicific/hedonic calculus)
- 5. Commitment to **equal consideration** - everyone's pleasure and pain counts equally
- 6. **Psychological egoism** – we act morally out of self-interest– all our ultimate motivations are self-interested.
- Bentham focused on making society work through socially imposed sanctions – he wanted to apply this idea to the laws of England. It helped him to justify laws for punishment, and to limit punishment (Bentham – punishment is in itself an evil, and can only be justified if it is a means to bring greater good)



Brief discussion

- **Q: Is it possible to be a psychological hedonist without being a psychological egoist? What would this entail?** When considering your answer, remember that these are theories (e.g. a psychological egoist is someone who *subscribes* to the theory of psychological egoism – this doesn't necessarily say anything about the person's own behaviour, or whether they are selfish or not)
- A: It is possible. This would be someone who believes that everyone does act out of pleasure all the time (psychological hedonism), but who gets pleasure from satisfying other people's desires or needs. **John Stuart Mill is an example of this.**



John Stuart Mill

John Stuart Mill refined Bentham's utilitarianism and today Classical Act Utilitarianism is primarily associated with Mill

- 1. Unlike Bentham, Mill rejected the idea that the mere quantification of utility was all that mattered. There are *also* differences of quality in utility, he says. "It is better to be a human being dissatisfied than a pig satisfied, better Socrates dissatisfied than a fool satisfied" Mill
- 2. Mill accepts psychological hedonism, but rejects Bentham's psychological egoism. Although he believed the only ultimate motivation of each human being is always to obtain pleasure/happiness, he thought that a person can obtain happiness through genuine moral motivations. There are also internal sanctions (governed by emotions): i.e. self-esteem, guilt, and conscience. These internal sanctions can be formed by education and social environment (upbringing, laws, ...)
- 3. Mill suggested a proof for the principle of utility.
- 4. Following social rules generally suffices (the principle of utility was at work in producing them, so they typically get it right). If there's conflict between rules, use the more fundamental principle of utility to determine what to do.

Summary of Mill's classical act utilitarianism

- 1. **Psychological hedonism.** Highlights the fundamental role of pleasure and pain in human life – the two “sovereign masters” as the only intrinsic motivations
- 2. **Normative Hedonism** – The only intrinsic good is pleasure - Equates the good with pleasure and the bad with pain.
- 3. We should approve or disapprove of an action on the basis of the amount of **utility** (pleasure/pain) it produces as a consequence – it is a **consequentialist** normative theory.
- 4. **Utility can be aggregated** (you can estimate how useful an action is overall), but utility also has quality (rejection of Bentham's quantitative-only util.)
- 5. Commitment to **equal consideration** - everyone's pleasure and pain counts equally
- 6. **Rejection of psychological egoism** – genuine moral motivations (produced by education) are internal emotional sanctions supporting moral action. A moral person has pleasure acting morally.

Mill's Greatest Happiness Principle

- “The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that 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.”
Mill, in *Utilitarianism*
- Correct interpretation: An action is right if and only if there is no other available action the agent could have done instead that has higher utility
- Utility = net of Pleasure – Pain
- In other words, **we need to maximize utility** (more accurate than speaking of maximizing pleasure – as the available action with most utility may involve pain)

Mill – the competent judges

- Mill: Bentham was wrong, it's not just about the quantity of pleasure
- “If I am asked what I mean by difference of quality in pleasures, or what makes one pleasure more valuable than another, merely as a pleasure, except its being greater in amount, there is but one possible answer.
- **Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure.**
- **If one of the two is, by those who are competently acquainted with both,** placed so far above the other that they prefer it, even though knowing it to be attended with a greater amount of discontent, and would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a **superiority in quality** so far outweighing quantity as to render it, in comparison, of small account.”

Another contribution – A proof of the principle of utility

- Bentham had said that the principle of utility “does not admit of direct proof”. Mill ambitiously thought he could provide this proof!
- “The only proof capable of being given that an object is visible is that people actually see it.... In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable is that people do actually desire it.
- If the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so.” (Mill, U, 81)

Clarification: Act Utilitarianism applies **only** to concrete actions

- Distinction: concrete (Paul steals an apple) vs. generic action (stealing)
- Think of the utility principle as a principle about the normative status of concrete actions only. It does not try to generalize about stealing, killing, and so on.

Activity in class/at home (time allowing)



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



Does utilitarianism capture all
that is relevant to making
moral decisions?



If not, what else matters?

- Jim finds himself in the central square of a small South American town. Tied up against the wall are a row of twenty Indians, most terrified, a few defiant, in front of several armed men in uniform. A heavy man in a sweat-stained khaki shirt turns out to be the captain in charge and, after a good deal of questioning of Jim that established he got there by accident while on a botanical expedition, explains that the Indians are a random group of the inhabitants who, after recent acts of protest against the government, are just about to be killed to remind other possible protesters of the advantages of not protesting.
- However, since Jim is an honoured visitor from another land, the captain is happy to offer him a guest's privilege of killing one of the Indians himself. If Jim accepts, as a special mark of the occasion, the other Indians will be let off. Of course, if Jim refuses, there is no special occasion and Pedro will do what he was about to do when Jim arrived, kill them all.
- Jim, with some desperate recollection of schoolboy fiction, wonders if he got hold of a gun, could he hold the captain and the rest of the soldiers to threat. It is quite clear from the setup that nothing of that kind is going to work — any attempt at that sort of thing will mean all the Indians will be killed, and himself. The men against the wall, and the other villagers, understand the situation and are obviously begging him to accept. **What should he do?**

Activity

Part I

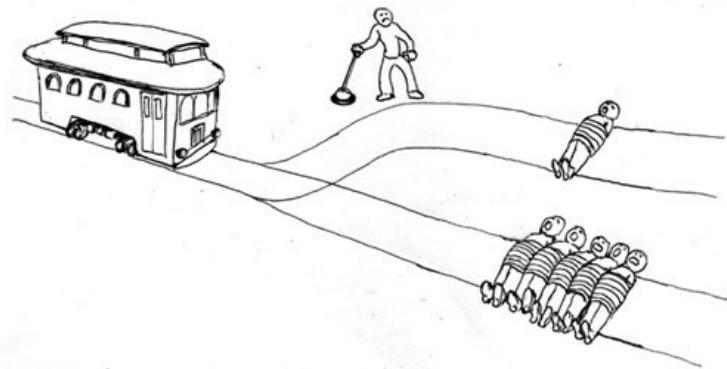
What to do if you were in Jim's shoes? Most importantly, why?

Is Jim responsible for the deaths of 20 if he refuses to kill anyone?



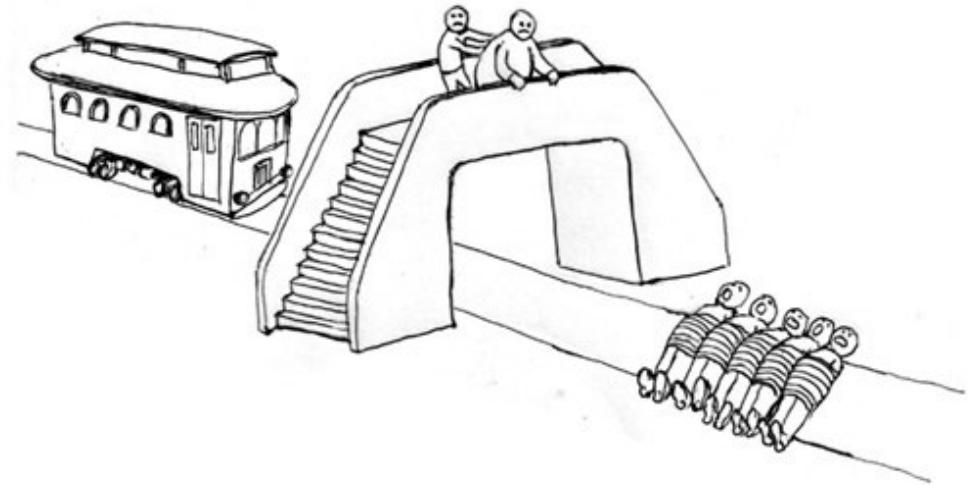
Trolley problem

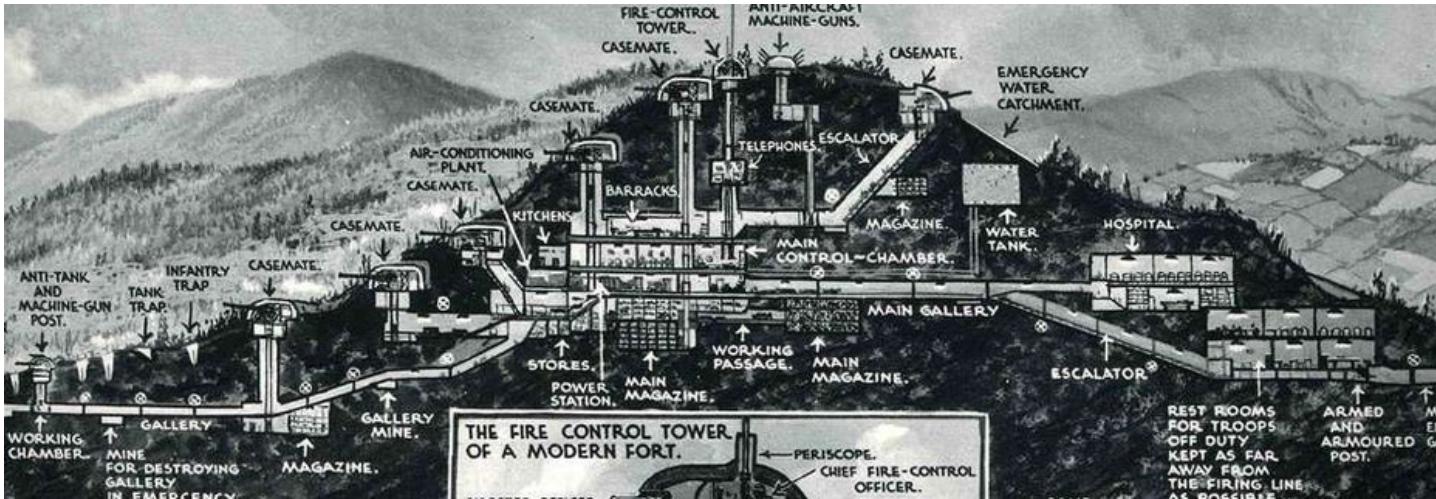
- Part II



Fat man variant

(Philippa Foot)





A lucky grey hat hacker

- Mr. Strangelove, a grey hat hacker, breaks into a server connected to Cheyenne Mountain, a NORAD facility (North American Aerospace Defense Command). He wants to show it's not sufficiently secure. In the process of showing how much control he can have he accidentally launches an ICBM toward Russia. The Russians are notified and through sheer luck manage to destroy the incoming missile.
- Outcome - Mr. Strangelove gets hired by NORAD and makes it more secure than ever. Communication with the Russians is improved and the world is now safer.

Does utilitarianism fail to capture all that is morally relevant?

- Is it moral to push the fat man to save five people?
- Is it moral to push your best friend to save five known war criminals?
- Did Mr. Strangelove act morally, although he almost started WWIII?
- Andy, a co-worker and good friend seriously messed up hacking a server as a challenge based on having inside knowledge from a client. Some damage he caused could be potentially traced back to him because very few had this knowledge. At some point you lie to your boss to protect Andy. You get lucky and nobody ever catches him, and it all goes well. Crisis averted, friendship is stronger than ever!
- According to utilitarianism, these actions are not just advised – they are morally required.