Lec 5

Objections to utilitarianism
Example: Dark Patterns

第5讲

对功利主义的反对意见
示例：黑暗模式

Dark patterns - Manipulative UX (User Experience) design

- Considered unfair and deceptive by the Federal Trade Commission (FTC) in the US in Oct 2021, by the EU in March 2022

- In Canada, there is no regulation explicitly addressing dark patterns, but we're going in that direction. Always keep in mind that the existence of a **legal loophole** doesn't mean an action is ethical.

- Existing privacy regulation already has some teeth so there's no guarantee that you and/or your company wouldn't face legal charges.

黑暗模式——操纵性的用户体验（UX）设计

- 2021年10月被美国联邦贸易委员会（FTC）认定为不公平且具有欺骗性，2022年3月被欧盟认定为不公平且具有欺骗性

- 在加拿大，目前尚无明确针对黑暗模式的监管法规，但我们正朝着这一方向迈进。请始终牢记：存在<code>**法律漏洞**</code>并不意味着某种行为就是合乎道德的。

- 现行的隐私监管法规已具备一定约束力，因此无法保证您和/或您的公司不会面临法律追责。

Personal Information Protection and Electronic Documents Act (PIPEDA)

What there is in Canada

- The Personal Information Protection and Electronics Document ACT (PIPEDA) relates to personal information – developed in the 90s. Enforcement of the act by the Office of the Privacy Commissioner of Canada has guidelines for obtaining meaningful consent regarding the collection, use and disclosure of personal information
- Canada's Anti-Spam Legislation (CASL) prohibits false and misleading statements in electronic messages
- If Bill C-27 got passed, the new Consumer Protection Act will have explicit language
- Section 16 of Bill C-27: "An organization must not obtain or attempt to obtain an individual's consent by providing false or misleading information or using deceptive or misleading practices. Any consent obtained under those circumstances is invalid."
- 6th Jan 2025 - Bill C-27 was dropped due to prorogation (interruption) of Parliament

Personal Information Protection and Electronic Documents Act (PIPEDA)

加拿大现行的相关法规

- 《个人信息保护与电子文件法》（PIPEDA）适用于个人数据——该法于20世纪90年代制定。加拿大隐私专员办公室负责执行该法，并就如何就个人数据的收集、使用及披露获取有效同意发布了相关指引
- 《加拿大反垃圾邮件法》（CASL）禁止在电子消息中作出虚假或误导性陈述
- 若C-27号法案获得通过，新《消费者保护法》将包含明确条款
- 《C-27号法案》第16条："组织不得获取或试图获取个人的同意，亦不得……" 信息，p或采用欺骗性或误导性 手段。在此类情况下所获得的任何同意均属无效。"
- 61月2025日——由于议会休会（中断），C-27号法案被撤回

Manipulating you to do what is best
Consent and avuncular ethics

- Utilitarianism is perhaps particularly apt to examine an issue like dark patterns as the theory is sensitive to context – rejects absolutism
- Kantian ethics, as we will see, would reject all manipulation regardless of intention, wouldn't consider context, and the outcome would be irrelevant
- Utilitarian ethics could possibly accept manipulation depending on the outcome
- Examples:
- 1. Trying to get someone to see as "default" something that is good for them. Declining contributing to pension plan is clear and possible but hard work – avuncular ethics
- 2. Visual design - If the user keeps clicking yes without reading, they get much needed insurance for a dangerous activity. Declining insurance requires spotting and checking a tiny box – more paternalistic and violates consent more

操纵你去做最正确的事：知情同意与
类叔伯式伦理

- 功利主义是　　perhaps p具体地　y ap去审视诸如'黑暗模式'之类的问题该理论对语境高度敏感，因而拒绝绝对主义
- 正如我们将看到的，康德伦理学将无条件拒斥一切形式的操纵，不考虑具体语境，且结果如何亦无关紧要
- 功利主义伦理学则可能依据最终结果而接受某种操纵

- 示例：
- 1. 试图让他人将某种对其有益的事物视为"默认选项"。拒绝参与养老金计划虽清晰可行，却需付出较多努力——这属于叔伯式伦理（avuncular ethics）。

- 2. 视觉设计——若用户在未阅读的情况下持续点击"是"，便会自动获得一项高风险活动所亟需的保险。而拒绝该保险则需用户主动发现并勾选一个极小的复选框——这种做法更具家长式作风，也更严重地违背了知情同意原则。

Revisions
Some typical issues

- <mark>No consideration of objections</mark>
- Summarizing utilitarianism down to one sentence or two (e.g. "it's about aiming for the general good") does almost nothing to demonstrate knowledge of the theory.
- Quotes on utilitarianism (Bentham, Mill) unaccompanied by an explanation in your own words typically do a very poor job of explaining the theory (see next slide).
- Complex calculations of utility (X causes pleasure, Y causes pain, Z would cause even more pleasure, …) do little to show knowledge of the theory. This just shows your knowledge of basic math!
- *Critical* discussion of problematic features (such as the ones we cover today) does a lot more to show knowledge
- Engage with complex issues successfully
- The approach of avoiding mistakes by only dealing with the basics guarantees a low grade

修订记录
一些典型问题

- 未考虑反对意见
- 将功利主义概括为一两句话 (e.g。"其核心在于目标" g 为了 g公共 '良善')几乎无法证明任何问题 对理论的理解。
- 仅引用功利主义相关名言（如边沁、密尔的论述），而未以自己的语言加以阐释，通常难以有效阐明该理论（参见下一张幻灯片）。
- 对效用进行复杂计算（例如：X 引发快感，Y 引发痛苦，Z 则会引发更强烈的快感，……）几乎无法体现你对该理论的理解。这仅仅表明你掌握了基础数学知识！
- 批判性探讨其存在问题的特征（例如我们今天所讨论的这些方面），更能体现你对该理论的深入理解。

- 成功应对复杂议题
- 仅局限于基础知识以避免出错的做法，注定只能获得低分。

Revisions - Misinterpreting act utilitarianism

- **The student writes: "We should maximize utility for the majority". Is this correct?**
- No
- In 1776, in his A Fragment on Government, Jeremy Bentham invoked what he described as a "fundamental axiom, it is the ==*greatest happiness of the greatest number*== that is the measure of right and wrong"
- 1. Clarification: utilitarianism in fact does **not** say we must serve the majority. It just so happens that in many cases (but not all) utility adds up when high numbers of people are found (e.g. kill 1 to save 19)
- *2.* Also, utilitarianism doesn't recommend uniform and wider distribution of utility so we can serve a "greater number'
- **What matters is only total *aggregated* utility. Choose that option no matter how many are helped, and how distribution looks**

修订内容——对行为功利主义的误解

- **学生写道："我们应当使多数人的效用最大化。"这种说法正确吗?**
- No
- 1776年，杰里米·边沁在其《政府片论》中指出：杰里米·边沁援引了他所称的"一条基本公理"，即"最大多数人的最大幸福，乃判定对错之标准"
- 1. 澄清：功利主义实际上并**不**主张我们必须服务于多数人。只是在许多（但并非所有）情况下，当涉及大量人群时，效用总和恰好更高（例如：杀死1人以拯救19人）
- *2.* 此外，功利主义并不主张通过均等化或扩大效用分配范围，以服务于"更多人数"
- **真正重要的是效用的总体加总值。无论受益人数多少、分配情况如何，都应选择该选项。**

Revisions - Misrepresenting act utilitarianism

- **The student writes: "it's moral because the benefits outweigh the costs"**
- This is an inaccurate summary of the actual procedure
- 1. You may have several options where the benefits outweigh the costs
- 2. You may have no options where this happens
- Student: You tell us not to use all these wrong definitions. What's the right one??

Utilitarianism is not a cost-benefit analysis, which has a *point of view*. What costs and benefits ***for X***?

Cost Benefit Analysis

With Current Gas Tank	With Safety Alteration
180 burn deaths	Cost = $11 per vehicle
180 serious burns	Total = $137 million
2100 Pintos burned	
Costs = $200 000 per death	
$67 000 per serious injury	Second alternative = Rubber Bladder
$700 per car	Cost = $5.08 per vehicle
Total = $49.5 million	Total ~= $64 million

Utilitarianism is also not about causing more good than bad

修订内容——
歪曲事实行为
功利主义

- **该学生写道："这是在道德上成立，因为其收益大于其成本"**
- 这是对实际程序的不准确概括
- 1. 您可能有若干种选择，其收益大于成本
- 2. 您可能没有任何一种选择能满足这一条件
- 学生：您告诉我们不要全部使用这些错误的（方法）定义。哪一个才是正确的？？

功利主义并非成本—收益分析，后者具有某种立场。那么，对X而言，哪些是成本，哪些又是收益对***X***?

Cost Benefit Analysis

| With Current Gas Tank	With Safety Alteration
180 burn deaths	Cost = $11 per vehicle
180 serious burns	Total = $137 million
2100 Pintos burned	
Costs = $200 000 per death	
$67 000 per serious injury	Second alternative = Rubber Bladder
$700 per car	Cost = $5.08 per vehicle
Total = $49.5 million	Total ~= $64 million

功利主义同样并非指"带来的益处多于损害"
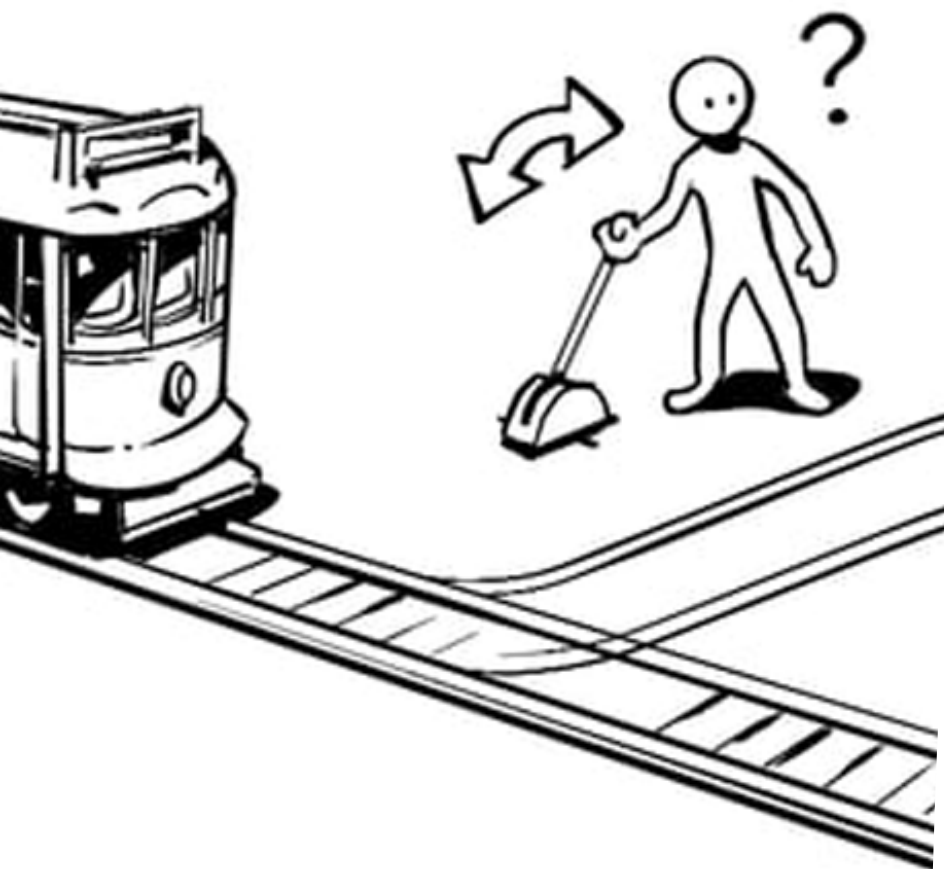
Using/Applying better than any copy-paste definition

We never ask for super short single-sentence definitions. Utilitarianism, like any other fairly complex theory, cannot be summarized in a single sentence. Use the *whole essay* to show your knowledge.

The closest and **still incomplete** "short" definition will perhaps be to describe the "algorithm" and *implement* it:

1) identify all affected by available actions

2) for each available action, aggregate all relevant utility (since the theory subscribes to normative hedonism, this is the sum of pleasure and pain, using ECI) considering the *actual* outcome and

3) choose the option that maximizes utility
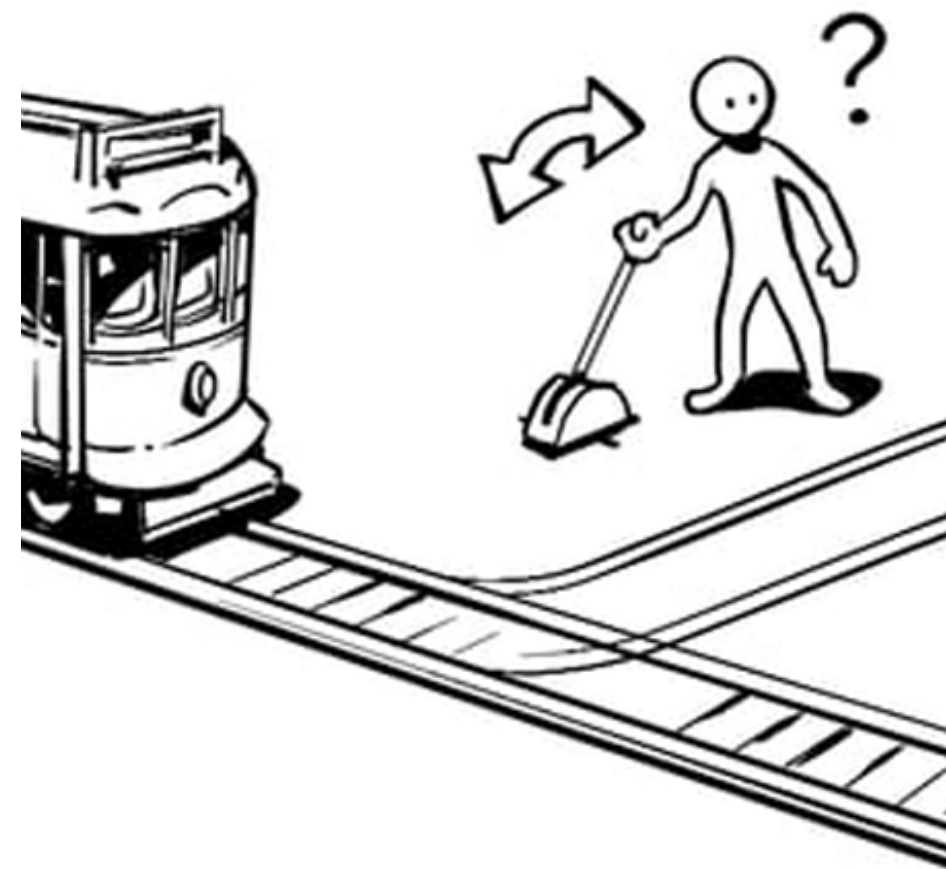
运用／应用：优于任何照搬照抄的定义

我们从不要求极简短的单句定义。功利主义，如同其他任何较为复杂的理论一样，无法用一句话概括。请使用整篇论文来展现你的理解。

最贴近且仍不完整的"简短"定义，或许可表述为：描述该"算法"并实现它：

1) 识别所有受可选行动影响的个体

2) 针对每一项可选行动，综合评估所有相关效用（由于该理论信奉规范性享乐主义，此处效用即为快乐与痛苦之总和，并采用经验效用指数（ECI）进行计算），同时考虑实际结果以及

3) 选择能使效用最大化的选项

Revisions – Why is it called *ACT* utilitarianism?

- **Suppose that someone says: "Respecting consent maximizes utility (as opposed to never respecting it), therefore we must always respect consent". Is this what act utilitarianism says?**
- No
- Clarification: *act (action)* utilitarianism evaluates actions. Respecting (all) forms of privacy is not an action (i.e. it's a *class* of actions). Act utilitarianism is not absolutist and evaluates each action on its own merits
- This would be rule utilitarianism – choose the *rule* (not action) that maximizes utility
- Often accused of collapsing into act utilitarianism, as an even better rule for maximizing utility would be "respect consent unless in a specific case you can maximize utility through a breach of consent, such as using dark patterns"

修订说明——为何称之为*ACT*功利主义?

- **假设有人提出："尊重同意能实现效用最大化（相较于从不尊重同意），因此我们必须始终尊重同意。"这是否就是行为功利主义的观点?**
- No
- 澄清：行为（行动）功利主义评估的是具体行为。尊重（所有形式的）隐私本身并非一个具体行为（即它属于一类行为，是一种行为类型）。行为功利主义并非绝对主义，而是针对每一项具体行为，依据其自身价值进行独立评估。

- 这属于规则功利主义——选择能最大化效用的规则（而非具体行为）。
- 规则功利主义常被指责实际上退化为行为功利主义，因为若要最大化效用，一条更优的规则或许是："尊重知情同意，除非在特定情形下，可通过违背同意（例如使用黑暗模式）来实现效用最大化。"

Revisions - Misinterpreting act utilitarianism

- **The student says "The action is right because it causes happiness". Is this right?**
- "Actions are right **in proportion** as they tend to promote happiness" – Proportionality doctrine of the Greatest Happiness Principle, Mill
- You may infer from the quote that a) right actions always cause happiness, b) there are many right actions and c) an action is right if it increases happiness. All incorrect.
- Clarification:
- A)The action recommended by utilitarianism may cause pain
- B) *Increasing* utility is not enough if it's not *maximized*. It's the "greatest" happiness principle.

修订——对行为功利主义的误解

- **该学生称："这一行为是正确的，因为它带来了幸福。"这种说法正确吗？**
- "行为之正当性 **与其促进**……的程度成正比" p幸福之程度 pp——功利主义p比例性 γ 学说 最大幸福原则，密尔
- 你可从该引文中推断出：a）正确的行为总会带来幸福；b）存在多种正确的行为；c）只要某行为能增进幸福，它便是正确的。以上三项均不正确。

- 澄清说明：
- A）功利主义所推荐的行为可能引发痛苦

- B）增加效用本身并不足够，除非它被最大化。这正是"最大"幸福原则的要义。

First, the three objections to utilitarianism as discussed by Mill himself

1. Doctrine of swine

2. Too high for humanity

3. Lack of time

首先，密尔本人所讨论的功利主义三大反对意见

1. 猪的学说

2. 对人类而言要求过高

3. 缺乏时间

Objection to utilitarianism - Doctrine of swine

- "Pleasure and pain are our sovereign masters" -psychological hedonism
- We should maximize utility - normative hedonism
- Objection, as described by Mill: "To suppose that life has no higher end than pleasure – no better and nobler object of desire and pursuit – they designate as utterly mean and grovelling; as a doctrine worthy only of swine"
- The objection is based on utilitarianism's claim that pleasure is the only thing that has intrinsic value (all other things have instrumental value).

对功利主义的反驳——"猪的学说"

- "快乐与痛苦是我们至高无上的主宰"——心理享乐主义
- 我们应当使效用最大化——规范享乐主义

- 密尔所提出的反驳："倘若认为人生别无更高目的，唯以快乐为终极目标；别无更佳、更崇高的欲求与追求对象，这种观点便被斥为彻底卑微而庸俗；实乃一种仅配得上猪的学说。"
- 该反驳基于功利主义的核心主张，即快乐是唯一具有内在价值之物（其余一切事物仅具工具性价值）。

Mill's response

- Mill: "some kinds of pleasure are more desirable and more valuable than others" – so we must consider quality as well as quantity
- How to know which are more desirable? Ask those who are properly acquainted with both, says Mill (the "competent judges")
- Is this a good response?
- How do we know who is a competent judge (who is properly acquainted with the two pleasures to compare)?
- What if judges disagree?
- Doesn't the idea of quality possibly smuggle non-utilitarian forms of morality evaluation to utilitarianism?

密尔的回应
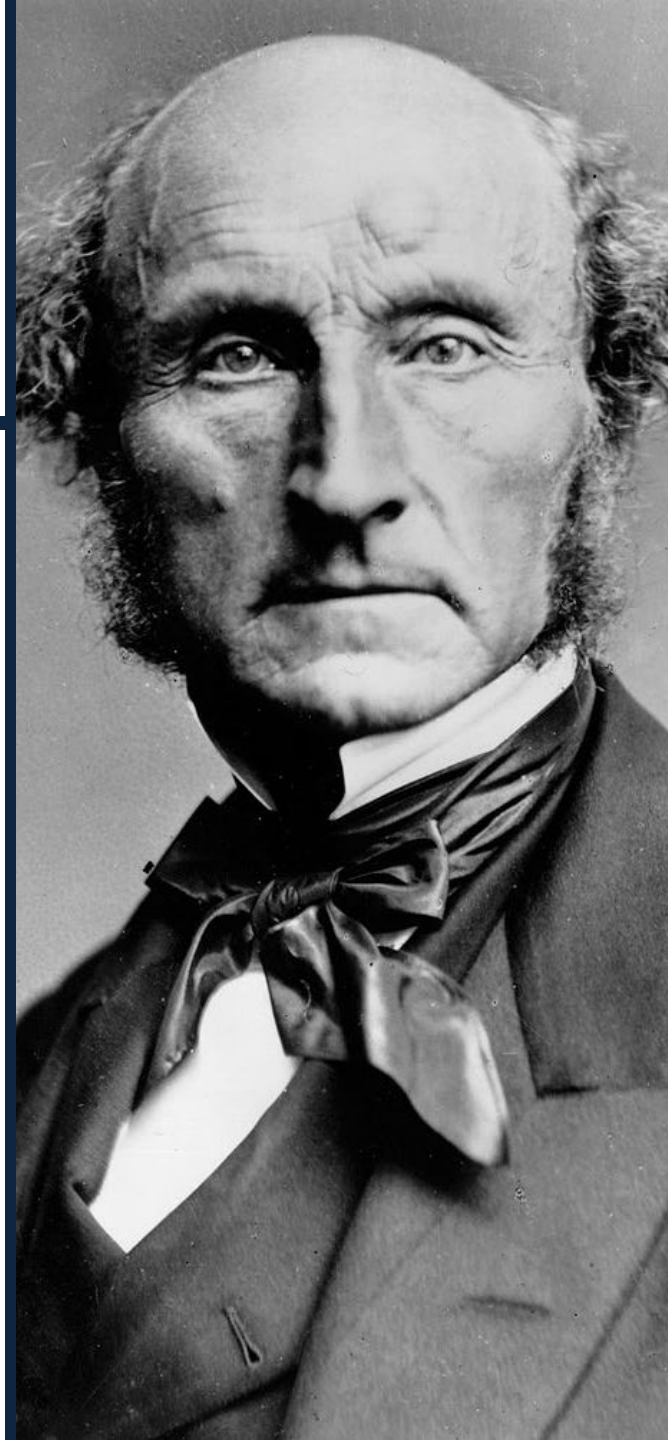
- 密尔："某些类型的快乐比其他类型更值得向往、更有价值"——因此，我们不仅需考量快乐的数量，还需考量其质量。
- 如何判断哪些快乐更值得追求？密尔指出，应询问那些对两者均有充分了解的人（即"有鉴别力的评判者"）
- 这一回应是否恰当？
- 我们如何确定谁是有鉴别力的评判者（即谁真正充分了解这两种快乐，从而能够加以比较）？
- 倘若这些评判者意见不一，又当如何？
- 将"质量"概念引入功利主义，是否可能暗中掺入了非功利主义的道德评价方式？
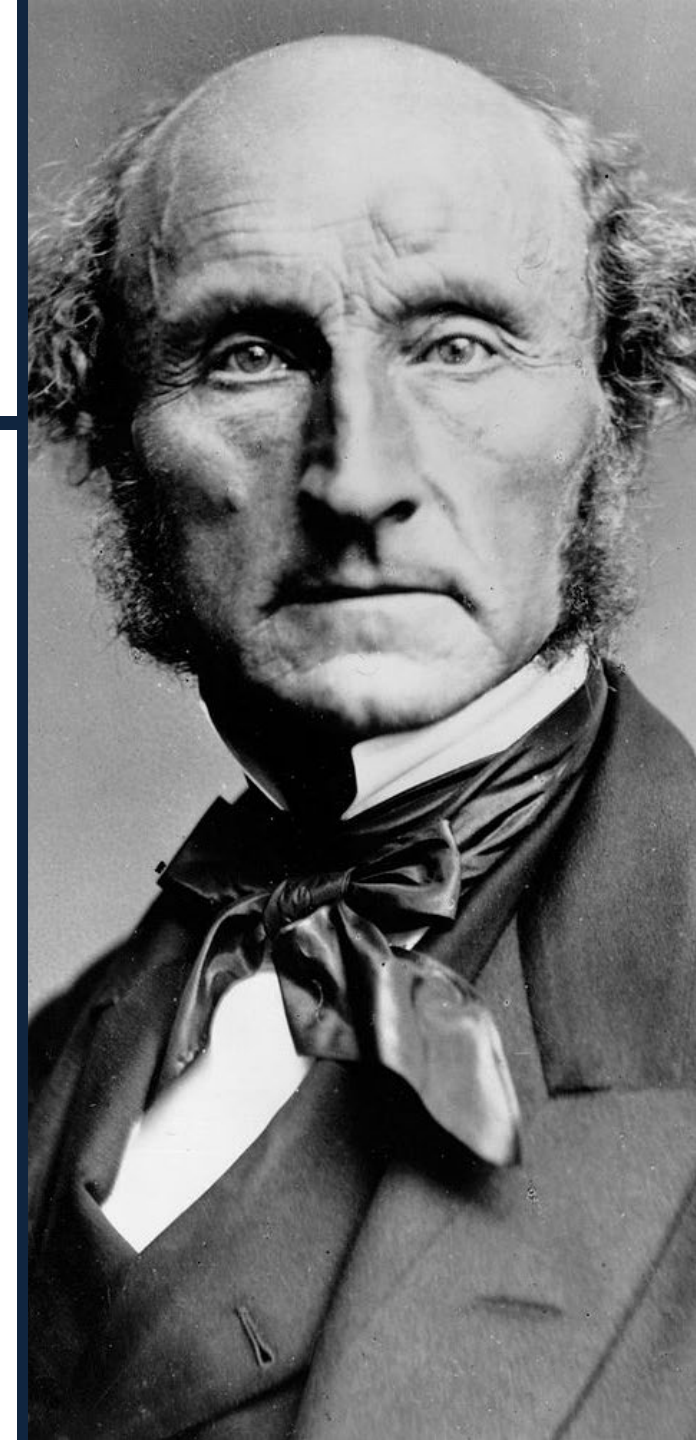
Objection to utilitarianism - Too high for humanity

- Mill's description of the objection - "It is exacting (demanding) too much to require that people shall always act from the *inducement* of promoting the general interests of society"

- Example of the objection: When we are kind to a neighbour, are we really **motivated** by the *general interest of society*? Surely, this cannot be the case. It is demanding too much.


14

对功利主义的反对——对人类而言要求过高

- 密尔对这一反对意见的描述—— "要求人们始终以促进社会普遍利益为动机来行动,未免要求过高。"

- 反对意见示例:当我们对邻居友善时,我们真的**是出于**对社会整体利益的考虑吗?这显然不可能。如此要求未免过于苛刻。

14

Mill's response

- Mill says that this is confusing the "rule of action with the motive of it. It is the business of ethics to tell us what are our duties, or by what test we may know them, but no system of ethics requires that the sole motive of all we do shall be a feeling of duty"

- But is it true that motives don't matter?
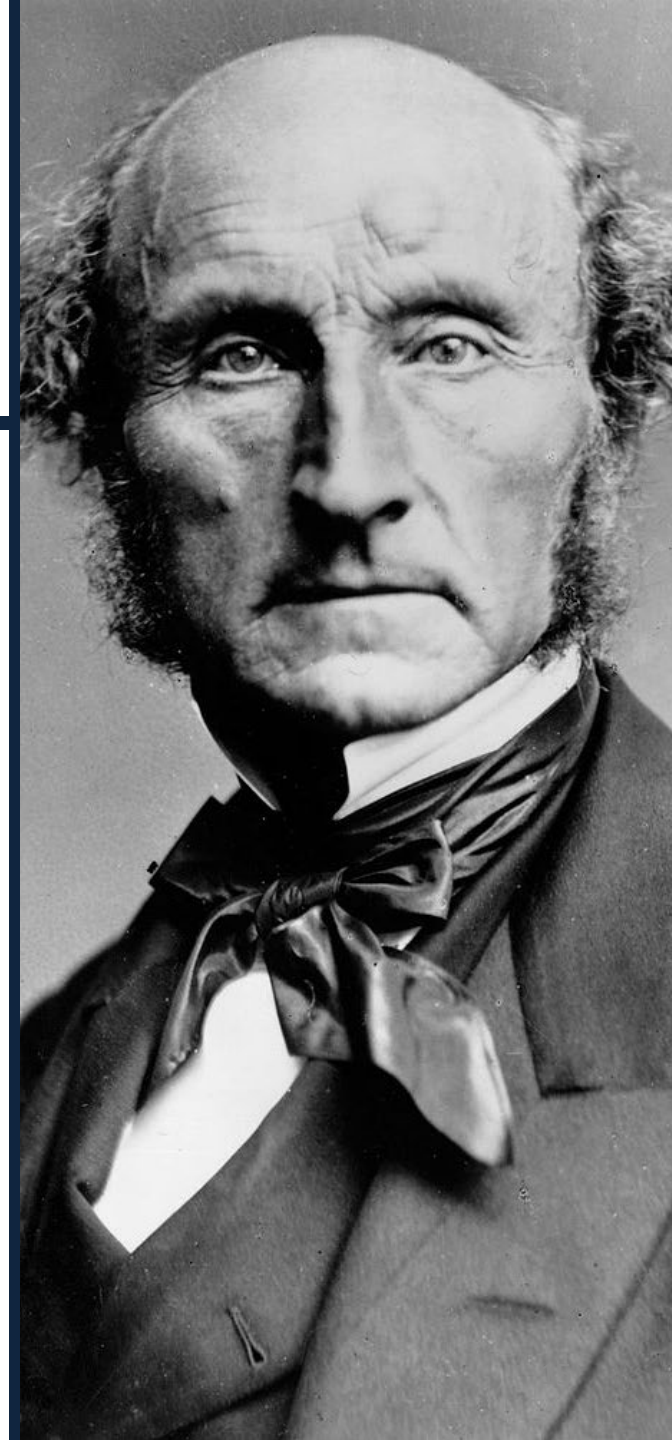
密尔的回应

- 密尔指出，这种观点混淆了"行为准则与行为动机"。伦理学的任务在于告诉我们何为义务，或通过何种标准来识别这些义务；但任何伦理体系均不要求我们将义务感作为一切行为的唯一动机"

- 但动机真的无关紧要?
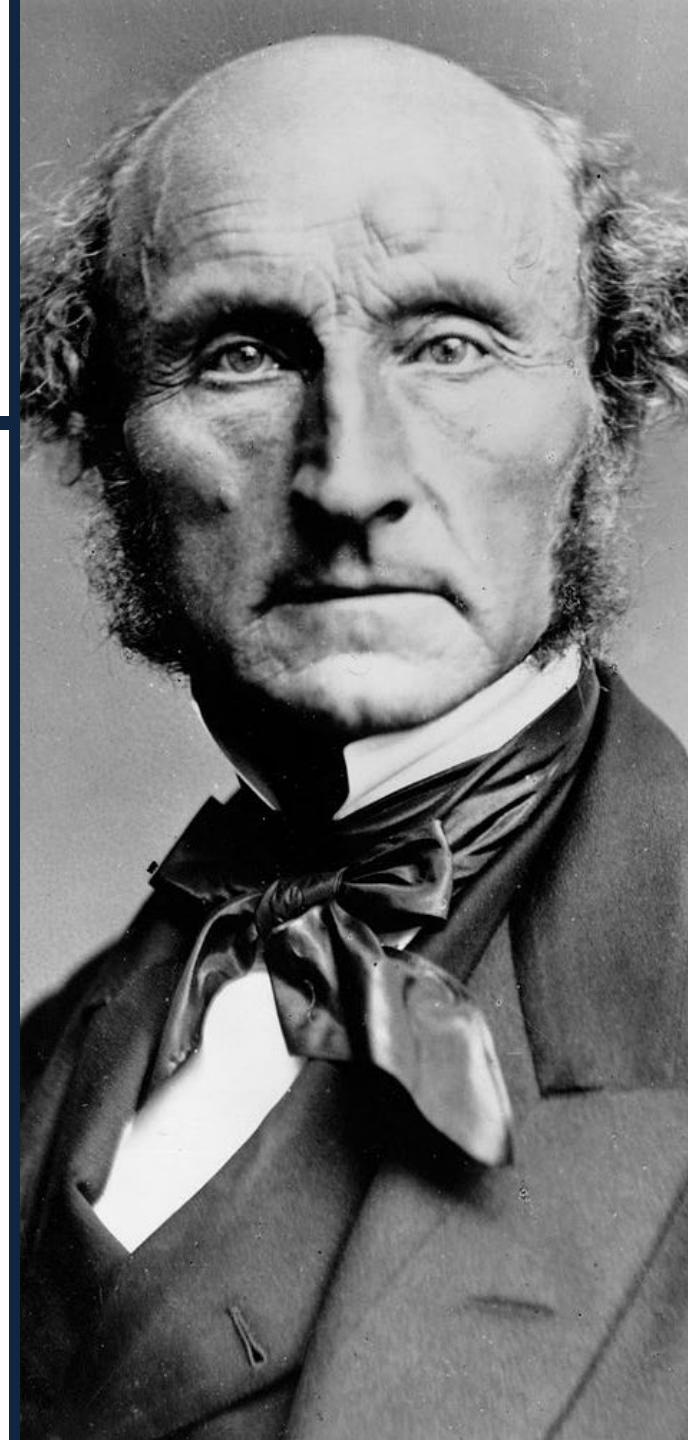
Is it true that motives don't matter in establishing right and wrong?

- Example from the Michael Sandel lecture: A doctor goes to a jungle to heal villagers but unknowingly she was carrying a disease and they all die
- Version B: A doctor goes to a jungle to wipe out all villagers. Knowingly she is carrying a disease she believes is lethal to them and they all die
- Version C: A doctor goes to a jungle to exterminate all villagers. Knowingly she carries a disease she believes is lethal to them, but they get immunity and are now healthier

在判定对错时，动机真的无关紧要吗?

- 迈克尔·桑德尔讲座中的一个例子：一名医生前往丛林为村民治病，却在无意中携带了一种疾病，导致所有村民死亡。
- 版本B：一名医生前往丛林，意图消灭所有村民。她明知自己携带着一种她认为对村民致命的疾病，最终导致所有村民死亡。
- 版本C：一名医生前往丛林，意图消灭所有村民。她明知自己携带着一种自认为会对村民致命的疾病，但村民却获得了免疫力，反而变得更加健康。

Objection to utilitarianism - Lack of time

- Mill's description of the objection: "There is not time, previous to action, for calculating and weighing the effects of any line of conduct on the general happiness"

对功利主义的 异议——时间不足

- 密尔对这一质疑的描述："在采取行动之前，并没有时间去计算和权衡任何行为方式对整体幸福所产生的影响。"

Mill's response

- "There is no difficulty in proving any ethical standard whatever to work ill if we suppose universal idiocy to be conjoined with it; but on any hypothesis after that, mankind must by this time have acquired positive beliefs as to the effects of some actions on their happiness" Mill

- Translation: any normative ethical theory can be accused to fail to provide correct answers if we assume the person using it is an idiot. Let's not assume that.
- One can know, by experience, that some actions (e.g. lying, corruption, torture) are likely to cause pain, other to cause pleasure. There is no need to calculate every time. Bentham had already made that point.
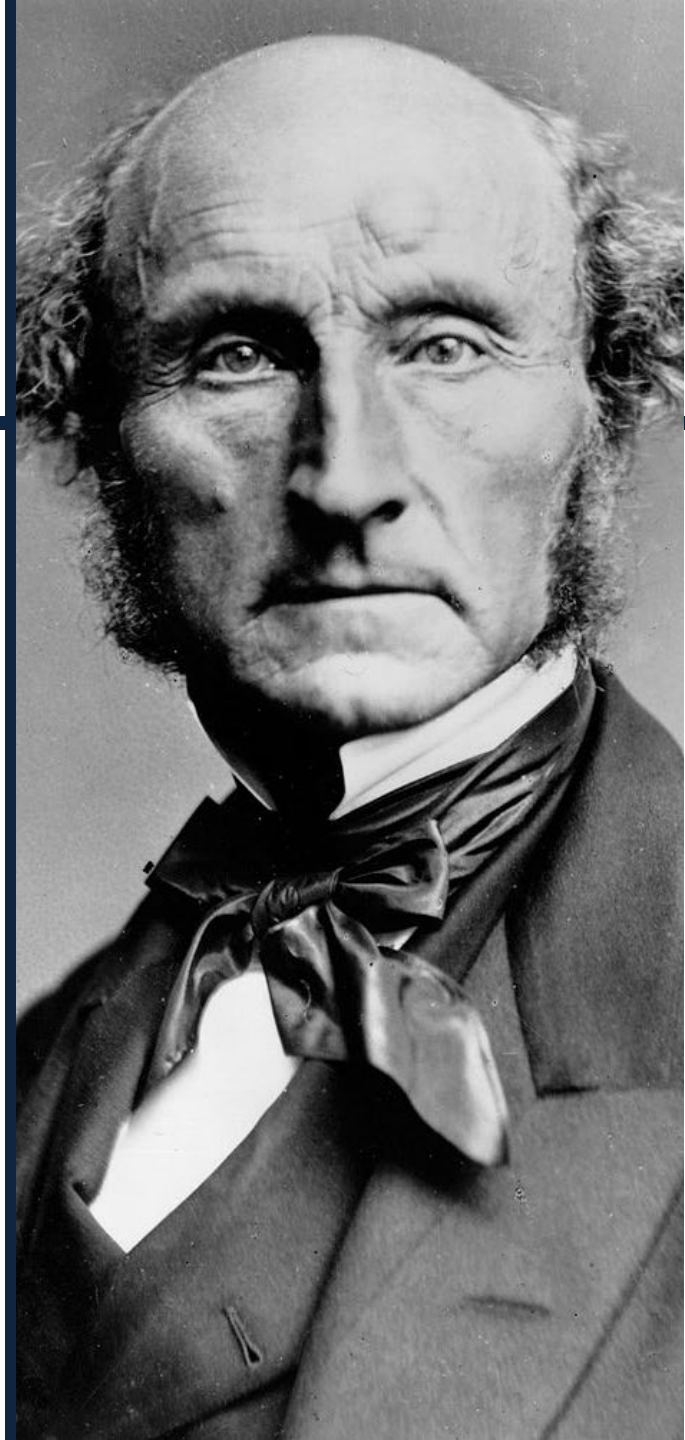
密尔的回应

- "若假设普遍愚昧与任何道德标准并存，则证明该标准会产生不良后果毫无困难；但在此假设之外的任何前提下，人类此时早已就某些行为对其幸福的影响形成了确定的信念。"——密尔

- 译文：任何规范性伦理理论，倘若假设其使用者是个白痴，都可被指责为无法给出正确答案。我们不应作此假设。
- 人们可通过经验得知，某些行为（例如说谎、腐败、酷刑）很可能引发痛苦，而另一些行为则可能带来愉悦。无需每次都进行计算。边沁早已指出过这一点。
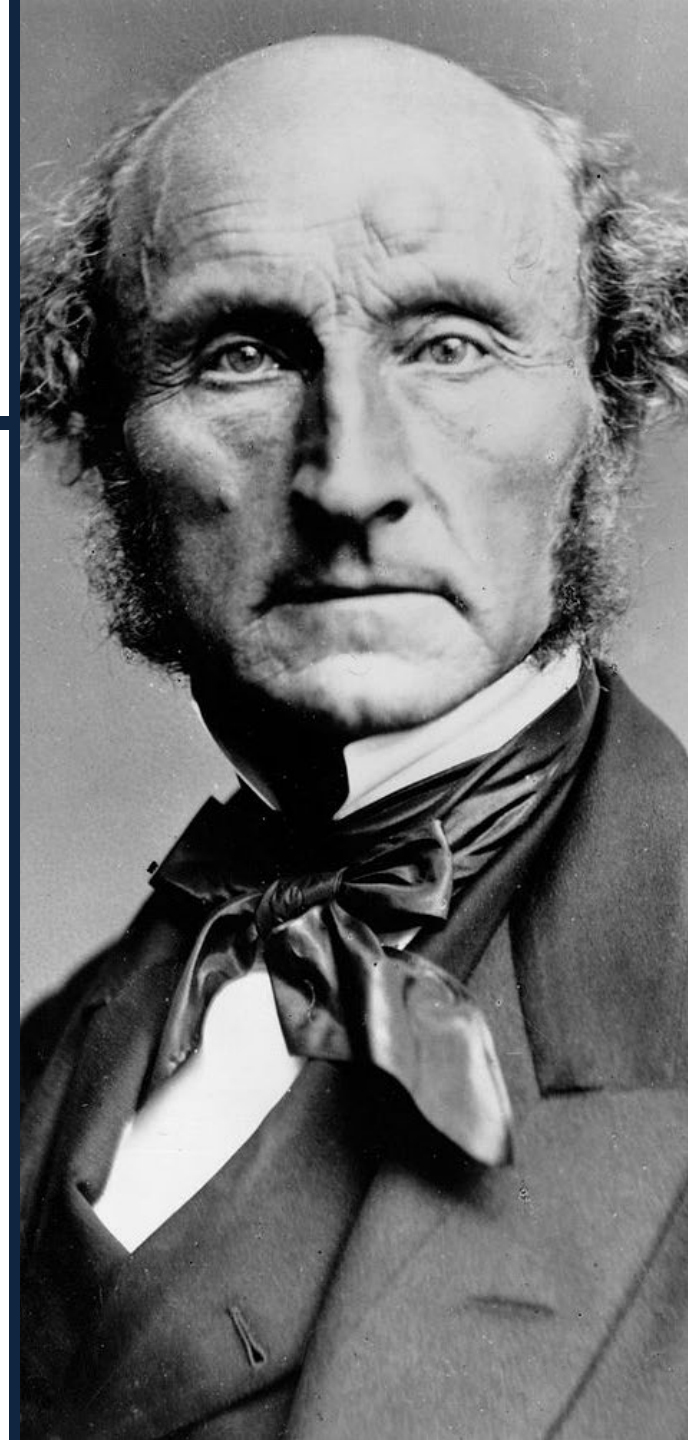
Additional problems with utilitarianism

- Motivational/psychological hedonism
- Normative hedonism
- Desire-fulfillment theory of welfare (Mill's proof)
- Aggregating utility
- Strong doctrine of negative responsibility
- Trivial actions
- Promises
- Punishment
- Justice
- Supererogatory action
- Impartiality

- 动机/心理享乐主义
- 规范性享乐主义
- 欲望满足论福利观（密尔的证明）

- 效用聚合工具
- 强烈的消极责任学说 责任
- 琐碎行为
- 承诺
- 惩罚
- 正义
- 超义务行为
- 公正无私

- Q: **Is it really true that all our actions can be ultimately explained by reference to pleasure?**
- A: Is my business/romantic partner cheating on me?
- Q: **Is it really true that all our actions can be ultimately justified by reference to pleasure?** (Is it true that we **should** act to maximize pleasure, because pleasure is the only good?)
- One famous argument/answer against both: Nozick's experience machine:
- If pleasure and pain are the *only* things that have ultimate value, you should want them even if they are not deserved/based on real accomplishments/etc.
- <mark>Marathon medal example</mark>

Psychological and Normative Hedonism

20

- 问：难道我们的一切行为最终真的都能以快乐为依据加以解释吗?
  - **难道我们的一切行为最终真的都能以快乐为依据加以解释吗?**
- 答：我的商业伙伴或恋爱对象是否正在背叛我?
- 问：**难道我们的一切行为最终真的都能获得充分的正当性证明吗?**
  by 指代　　　p快乐? (是否我们**应当**采取行动以最大化快乐，因为
  快乐是唯一的善?
- 一个著名的论证/回答反对两者：诺齐克的体验机器：
- 倘若快乐与痛苦是唯一具有终极价值的东西，那么即便它们并非源于应得的成就、真实功绩等，你也理应追求它们。
- 马拉松奖牌示例

心理与规范
享乐主义

20

The Experience Machine by Yescela Vorazan

Nozick's experience machine

- Anarchy, State and Utopia (1974)
- The experience machine:
- 1. Can give you any experiences you want
- 2. Perfectly simulates a life of achievement, pleasure, love, etc
- 3. It is indistinguishable from reality
- 4. You must plug in permanently (you won't remember choosing)

- **Do you plug in?**

诺齐克的经验机器

- 《无政府、国家与乌托邦》（1974年）
- 体验机器：
-     1. 可为你提供任何你想要的体验

- 2. 完美模拟一种充满成就、愉悦、爱情等的人生
- 3. 其体验与真实世界无法区分
- 4. 你必须永久接入该机器（接入后将不记得自己曾做过这一选择）

- **你是否已接入？**

Nozick: 3 reasons not to plug in

- *1. We want to* do *certain things, and not just have the experience of doing them.*
  - "It is only because we first want to do the actions that we want the experiences of doing them." (Nozick, 43)

  (in other words, it's not the experience – the pleasure – that matters the most. Applying to our example, we want the person to be well, and secondarily to have pleasure *because* they are doing well)
- *2. We want to* be *a certain sort of person.*
  - "Someone floating in a tank is an indeterminate blob." (Nozick, 43)

  (in other words, character matters to us, rather than just pleasure – whether we are truly a good person, or just falsely believe we are a good person)
- *3.* Plugging into an experience machine limits us to a man-made reality (it limits us to what we can make).
  - "There is no *actual* contact with any deeper reality, though the experience of it can be simulated." (Nozick, 43)

  (in other words, it seems truth matters to us - **ignorance may be bliss, but it's worthless bliss**)

- Conclusion: pleasure is not the ultimate motivation + pleasure is not the only thing that has intrinsic value.

诺齐克：拒绝接入体验机的三个理由

- 1．我们希望真正去做某些事情，而不仅仅是拥有做了这些事情的体验。
  - "我们之所以想要拥有做某事的体验，首先是因为我们本就想去做那些事情。"（诺齐克，第43页）

  （换言之，真正至关重要的并非体验本身——即快感。套用我们的例子来说，我们首先希望此人身体健康，其次才希望其因身体康健而获得快感 因为 他们状态良好）
- 2．我们希望成为某种特定类型的人。
  - "漂浮在水箱中的人只是一个不确定的团块。"（诺齐克，第43页）

  （换言之，品格对我们至关重要，而不仅仅在于快乐——关键在于我们是否真正是一个好人，而非仅仅错误地相信自己是个好人）
- *3.* 接入体验机将我们局限于人为构建的现实（即局限于我们所能创造之物）。
  - "我们并未与任何更深层的实在发生真实接触，尽管这种接触的体验可以被模拟。"（诺齐克，第43页）

  （换言之，真理似乎对我们至关重要——**无知或许带来幸福，但这种幸福毫无价值**）

- 结论：快乐并非终极动机；+ 快乐并非唯一具有内在价值之物。

Desire-fulfillment theory of welfare (Mill's proof)

- Mill – We desire pleasure, therefore it is desirable

- 2 interpretations for "desirable":
- We desire pleasure, therefore we desire pleasure – trivially true but circular
- We desire pleasure, therefore it is good – **clearly wrong**

- Mill argued each individual is better positioned than anyone else to know what is good for himself/herself. This is one argument he used to refuse interference from the state in personal matters that don't affect society.

- Objections:
- But don't humans desire all sorts of things that are bad for them?
- If I desire to be hurt simply because I have low self-esteem, is it desirable that I am hurt? No.
- If I desire crack because I'm addicted to crack, does that show that crack is desirable (good for me)? No.

福利的欲望满足理论（密尔的证明）

- 密尔 — 我们欲求快乐，因此快乐是可欲求的

- 2 介于 p "可欲求的" 一词的几种解释：
- 我们欲求快乐，因此我们欲求快乐——这虽属显然为真，却陷入循环论证
- 我们欲求快乐，因此快乐是善的——**明显错误**

- 密尔认为，每个个体都比任何他人更清楚什么对自己有益。他正是基于这一论点，拒绝国家对不涉及社会利益的个人事务进行干预。

- 反驳意见：
- 但人类难道不会渴望各种对自己有害的事物吗？
- 如果我仅仅因为自尊心低下而渴望受到伤害，那么我遭受伤害就真的可取吗？不。
- 如果我渴望快克是因为我对快克上瘾，这是否说明快克是值得向往的（对我有益）？不。

Aggregating utility
Brief discussion

- **Is it OK to aggregate all utility? Or must we respect the rights of individuals, even at a cost to general happiness?**
- 1. The use of dark patterns leads a user to accept to share photos from his phone that he wished were private.
- But as it turns out in one particular case, eventually the RCMP uses these photos to identify him as a sex offender and get a conviction.
- **Was the original violation of privacy ethical?**

效用聚合：简要讨论

- **将所有效用进行聚合是否合理？抑或我们必须尊重个体权利，即便这会以整体幸福为代价？**
- 1. 暗黑模式的使用诱导用户同意共享其手机中本希望保持私密的照片。
- 但事实证明，在某一特定案例中，加拿大皇家骑警最终利用这些照片将他识别为性犯罪者，并成功定罪。
- **最初的隐私侵犯行为**
  道德上的?

Strong doctrine of negative responsibility in Killing One to Save 19

- Bernard Williams asks whether we are really responsible for the things we *don't* do *as much* as we are responsible for the things that we do, asks Bernard Williams.
- Note that he's not denying that one can be guilty of negligence in certain contexts.
- Bernard Williams **integrity argument** – One may have vowed never to kill. Is not killing as blameworthy as killing? Perhaps an argument based on integrity has some purchase, thus placing the blame entirely or at least primarily on Pedro rather than oddly placing it on the person who refuses to hurt anyone out of moral integrity.
- Or imagine that after killing 1 to save 19, Pedro does release the 20, but then he gets another 20. Jim is now asked by psychopath Pedro to engage in increasingly horrible actions, again hurting 1 to save 19. The utilitarian will always make the same calculation, and will keep always doing what Pedro expects…

《杀死一人以拯救十九人》中的强式消极责任学说

- 伯纳德·威廉姆斯提出质疑：我们是否真如对自己所作所为负有责任那样，同样对我们未做之事负有责任？不做之事——伯纳德·威廉姆斯如此发问。

- 请注意，他并未否认在某些情境下，人们可能因疏忽而担责。

- 伯纳德·威廉姆斯**主 grity arg张**——某人　　y 可能曾立誓永不杀人。不杀人与杀人同样应受谴责吗？或许可基于道德操守提出某种论据，从而将全部或至少主要责任归于佩德罗，而非不合常理地将责任归于那位出于道德操守而拒绝伤害任何人的个体。

- 或者设想一下：佩德罗为救下19人而杀死1人后，确实释放了这20人；但随即他又抓到了另外20人。此时，精神变态者佩德罗要求吉姆实施愈发骇人听闻的行为——再次牺牲1人以拯救19人。功利主义者将始终进行同样的权衡计算，并持续做出佩德罗所期望的抉择……

Trivial actions

- According to utilitarianism, utility must be maximized.
- This seems to entail that we can't take time out. Ever.
- Can't we just do something that is trivial, when we are not concerned with maximizing utility (e.g. buying new shoes, watch Netflix, go for a walk)? Do we *always* need to be maximizing utility?

琐碎行为

- 根据功利主义，必须使效用最大化。

- 这似乎意味着我们永远无法暂停片刻。

- 当我们并不在意效用最大化时（例如买双新鞋、看网飞剧集、散个步），难道就不能做些简单 trivial 的事情吗？我们是否总是必须追求效用最大化？

How utilitarians see promises

- Promises will rarely be kept, since you only need to keep a promise when it maximizes utility
- When a promise is kept, there's no requirement to do it *because* it was promised.
- For utilitarianism, breaking a promise may cause **pain** (and that relevant for the theory), but otherwise there's no intrinsic moral obligation to keep promises

功利主义者如何看待承诺

- 承诺极少会被遵守，因为你只需在符合功利原则时才需要遵守承诺。当它能实现效用最大化时
- 当一项承诺被履行时，并非因其曾被许下而必须如此履行，而是因为这样做符合功利原则。
- 对功利主义而言，违背承诺可能引发**痛苦**（而这在该理论中具有相关性），但除此之外，并不存在恪守承诺的内在道德义务。

Example from the Sandel lecture:
Promise to the dying man

- Man and grandfather go sailing.
- Shipwrecked on an island, the grandfather is dying and asks to be buried and that his body is treated with respect. The grandson promises but later considers whether to use the body for fishing bait.

桑德尔讲座中的示例：对临终之人的承诺

- 男子与祖父一同出海航行。
- 祖父在孤岛上遭遇海难，生命垂危，临终前请求将自己安葬，并希望遗体得到尊重对待。孙子答应了这一请求，但后来却开始考虑是否该将祖父的遗体用作钓鱼饵料。
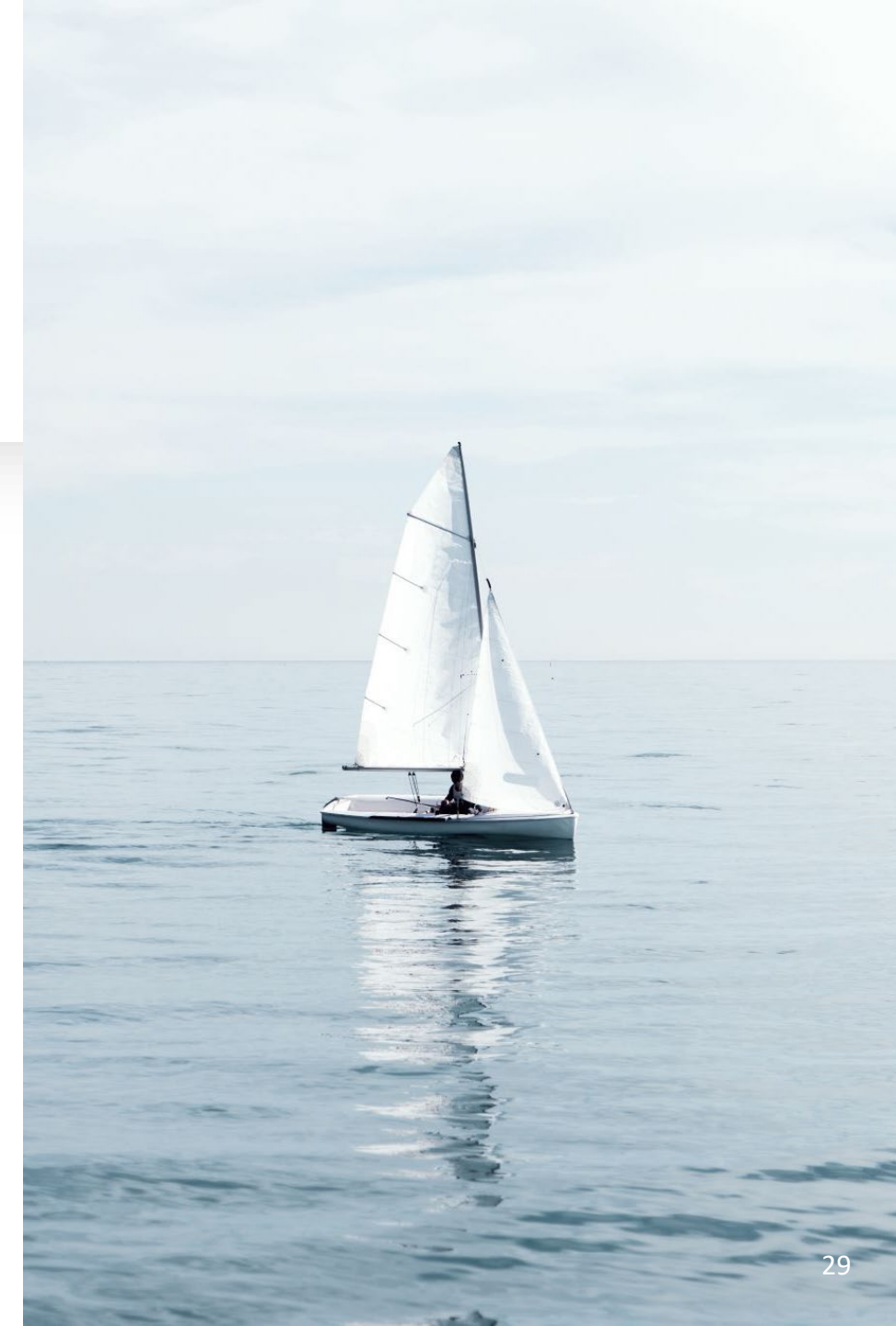
Utility

- Use body for bait   +25
- Bury body, according to promise   -10
- Let body rot where it is   -2
- Throw body in sea   -1

- Discussion
- Assume that the grandson will die if he treats the body "with respect".
- 1. Does he have a moral obligation to keep his promise?
- 2. Should he have promised in the first place?
- 3. Should the grandfather have asked him to make that promise, knowing the consequences for his grandson?

效用

- 以尸体为诱饵 +25
- 依承诺安葬遗体：-10
- 让尸体原地腐烂：-2
- 将尸体抛入大海：-1

- 讨论
- 假设孙子若以"尊重"的方式处理尸体，便会死亡。
- 1. 他是否有道德义务信守自己的承诺?
- 2. 他最初是否就不该做出这一承诺?
- 3. 祖父在明知这一承诺将给孙子带来后果的情况下，是否仍应要求孙子作出该承诺?

Crime and punishment

- Punishment produces utility through:
- Rehabilitation
- Deterrence
- Vengeance

罪与罚

- 惩罚产生效用
  通过：
- 康复
- 威慑
- 报复

Punishment - problems

- **Rehabilitation**
- Not seen as a right (good in itself) from the point of view of utilitarianism. If we happen to maximize utility by leaving the person e.g. in solitary confinement for life, so be it.
- **Deterrence**
- Why not frame an innocent and punish him in public? Or convict three people, to make sure the one guilty person is punished?
- **Vengeance**
- It can cause pleasure, so any punishment that gives pleasure to enough people is morally justified. Whether the person committed the crime or not is not intrinsically morally relevant for the theory.

- **All these problems derive from not seeing justice and a defense of human rights as values in themselves. Acting on them will have only derivative value when considering utility (e.g. protect someone's freedom of speech if doing that maximizes utility in one particular instance).**

- 康复
- 功利主义视角下，康复并不被视为一种权利（即其本身即为善）。倘若我们恰好能通过将某人——例如终身单独监禁——来实现效用最大化，那便如此行事。
- 威慑
- 那么，为何不捏造证据、公开惩罚一名无辜者？又或者，为何不判处三人有罪，以确保其中真正有罪的那个人受到惩处？

- Veng伊安斯
- 它可能带来快感，因此，任何能让足够多的人感到愉悦的惩罚，在道德上都是正当的。至于被惩罚者是否真正实施了犯罪，在该理论看来，并不具备内在的道德相关性。

- 所有这些问题都源于未能将正义与人权保障本身视为内在价值。若仅从效用角度出发（例如：仅在某一特定情形下，保护某人的言论自由能实现效用最大化），那么践行这些原则便只具有派生性价值。

Justice and fairness

- Q: Peter and Sarah worked equally as hard and equally as long. They get paid a single amount to split. Should they split it 50-50?
- A: For utilitarianism, this should be done if it maximizes utility. But if Peter is sad unless he gets 60% of the amount, this would factor in the calculation.

- Because in utilitarianism there is:
- Equal consideration of interests, but…
- Utility is aggregated

正义与公平

- 问：彼得和莎拉付出的辛劳程度与工作时长完全相同。他们共同获得一笔固定报酬，需自行分配。这笔钱是否应按五五开的方式平分?
- A：依据功利主义，若此举能实现效用最大化，则应当如此行事。但倘若彼得未获得总额的60%便会感到悲伤，这一因素也需纳入计算。

- 因为在功利主义中存在以下原则：
- 对各方利益给予同等考量，但……
- 效用是加总计算的

Impartiality and Supererogation – Singer's "Famine, Affluence, and Morality"

- Singer – Must we help Bengali refugees? What obligations do we in affluent nations have toward victims of natural disasters or other disasters of that magnitude?
- Singer's answer: We must be impartial (emotional proximity and physical distance do not matter to utilitarianism).
- Singer: It's also **not mere** *charity* to help (charity defined as beyond the call of duty – ethical, but not strictly *required*). It's our **duty** – definitely required!

- Note that these ideas can easily be applied to computing and are obviously not specific to natural disasters, refugees, etc:
- You're asked to implement a dark pattern that impacts users in India. How far you go to protect those users?
- Should you point out to your boss that this in fact a dark pattern? Does it depend on how much is at stake for you?

公正性与超义务性——辛格《饥荒、富裕与道德》

- 辛格：我们是否必须援助孟加拉国难民？我们这些富裕国家的民众，对遭受自然灾害或其他同等规模灾难的受害者负有何种道德义务？
- 辛格的回答是：我们必须秉持公正立场（情感上的亲近性与物理上的距离，在功利主义框架下均无关紧要）。

  辛格指出：援助他人也并非单纯的<code>慈善</code>行为（此处'慈善'被定义为超出义务范围的行为——虽具道德价值，却非严格意义上的义务）；
- 而是我们的<code>义务</code>——确属必须履行之责！

- 请注意，这些观点可轻易延伸至计算机领域，且其适用性显而易见。不仅限于自然灾害、难民等情况：
- 你被要求实施一种对印度用户造成负面影响的黑暗模式。你会为保护这些用户做到何种程度？
- 你是否应向你的上级指出，这实际上是一种黑暗模式？这是否取决于此事对你自身利害关系的大小？

Impartiality

- Singer: "[W]e cannot discriminate against someone merely because he is far away from us" (in distance or psychologically – friend or stranger)

- Discussion
- Three kids fall in the water. You can save two kids who are strangers or your own kid. Must you save the two strangers?
- What if it is was 11 strangers versus your family of 10 (partner, grandparents, parents, kids)? Is it your moral obligation to let your whole family die?
- A1. Absolutely. Impartiality is the only way to be moral.
- A2. Possible objection against utilitarianism: "No ought without can". Almost no one *can be* that impartial. So there's no obligation to be that impartial.
- A3. Possible objection against utilitarianism: It's actually wrong to be that impartial. There are special duties to those who rely on us (e.g. our children, partner, friends).

公正性

- 歌者："[W]e 不能仅仅因为某人离我们很远（无论是在物理距离上还是心理距离上——是朋友还是陌生人）就歧视他。"

- 讨论
- 三个孩子落水了。你可以救两名陌生的孩子，也可以救你自己的孩子。你是否必须选择救那两名陌生人？
- 如果换成11名陌生人与你由配偶、祖父母、父母和子女共10人组成的家庭呢？你是否有道德义务让你的整个家庭牺牲？
- A1. 当然如此。唯有秉持公正性，才能做到合乎道德。
- A2. 对功利主义的可能质疑："无能为力，便无义务"（No ought without can）。几乎无人`能够`做到如此公正无私，因此也就没有义务必须做到如此公正无私。
- A3. 对功利主义的可能质疑：事实上，做到如此公正无私反而是错误的。我们对依赖我们的人（例如子女、伴侣、朋友）负有特殊的道德义务。

Singer on supererogation – Actions beyond the call of duty

- "It is beyond the scope of my argument to consider whether the distinction [between duty and charity] should be **redrawn or abolished** altogether"
- In other words, Singer proposes to change or eliminate the threshold of supererogation. He seems to propose here the more reasonable redrawing:
- "The strong version, which required us to prevent bad things from happening unless in doing so we would be sacrificing something of comparable moral significance, does seem to require reducing ourselves to the level of marginal utility. **I should also say that the strong version seems to me to be the correct one. I proposed the more moderate version** - that we should prevent bad occurrences unless, to do so, we had to sacrifice something morally significant" Singer

辛格论超义务行为——超越职责要求的行为

- "我的论证范围并不涵盖如下问题：即义务与慈善之间的区分[是否应当]重新界定**或彻底废除**。"
- 换言之，辛格主张改变或取消超义务行为的界限。他在此处似乎提出了一种更为合理的方案：重新界定：
- "强版本要求我们防止坏事发生，除非在这样做时，我们将牺牲具有同等道德重要性的东西；该版本确实似乎要求我们将自身降低至边际效用水平。**我还应指出，就我而言，强版本似乎是正确的。我所提出的是一种较为温和的版本**——即我们应当防止坏事发生，除非为做到这一点，我们必须牺牲某种具有道德重要性的东西。重要的" 歌手

The threshold of supererogation

- If it's **redrawn**, then we need to speak of a moral duty to do *some* charity
- If it's **abolished**, no duty is considered supererogatory (above and beyond the call of duty) – note that this what classical act utilitarianism actually calls for!

- A consequence of abolishing the threshold, which Singer and classical act utilitarianism say is ultimately right – we can no longer distinguish between:

- **Duty** (it is out duty to give - ethically mandatory)
- **charity** (usually considered to be within the realm of generosity above the call of duty – moral to give, but optional)

超义务行为的阈值

- 倘若该图被**重新绘制**，那么我们就必须谈及一种道德义务，即开展某些慈善活动
- 倘若该阈值被**废除**，则没有任何义务会被视为超义务行为（即超出职责要求的行为）——需注意，这恰恰是经典行为功利主义所主张的观点！

- 废除该阈值所带来的一项后果（辛格与经典行为功利主义均认为此举最终是正确的）在于：我们再也无法区分以下两类行为——

- **义务**（给予是我们的义务——在伦理上具有强制性）
- 慈善行为（通常被视为超越义务范畴的慷慨之举——给予是道德上的应然之举，但并非强制性义务）

Threshold of supererogation
Do we move it up (moderate proposal) or eliminate it (utilitarianism proper)?

超 越 动作
我们是将其提升（温和型方案），还是彻底取消（严格意义上的功利主义）？

Strong proposal – always maximize utility
(actual utilitarianism)

强有力主张——始终最大化效用（实际功利主义）

Charity (moral but optional)

慈善（道德层面——但属自愿）

Threshold of superogation

超义务行为的阈值

Duty (moral - must be done)

义务（道德层面——必须履行）

Discussion (time permitting) Or consider this at home

If you still want to be a utilitarian, you should have responses to the objections raised.

1. Consider the objections we discussed (see a list on the next slide)
2. Do you have responses for each? Identify the most problematic objection, and come up with an example to illustrate how its problematic
3. In light of your responses, ask yourself whether you are a utilitarian?
4. Can we entirely give up on considerations of social utility? (given enough social utility, does it ever tip the scale, where we should be prepared to engage in some actions even if they seem to violate individual rights?)

讨论 （时间 允许的情况下） 或者考虑 这一点在霍 me

倘若你仍希望成为一名功利主义者，就应当对所提出的各种反对意见作出回应。

1. 请考虑我们之前讨论过的各种反对意见（参见下一张幻灯片中的列表）
2. 您是否对每个问题都准备了回应？找出最成问题的反对意见，并构思一个示例 ，以说明其为何成问题
3. 结合您的回应，请扪心自问：您是否是一名功利主义者？
4. 我们能否完全放弃社会效用方面的考量？（在社会效用足够充分的情况下，它是否总能成为决定性因素，使我们愿意采取某些行动，即便这些行动看似违背了个体权利）权利？）

Popular objections to utilitarianism

- It's a doctrine of swine
- Lack of time to calculate utility
- Too much for humanity
- Motivational/psychological hedonism seems to be mistaken
- Normative hedonism seems to be mistaken
- Desire-fulfillment theory of welfare (Mill's proof) doesn't work
- Aggregating utility often leads to counter-intuitive results
- Strong doctrine of negative responsibility is mistaken
- All sorts of trivial actions are incorrectly classified as morally wrong
- Keeping promises seems of little importance
- Punishment doesn't need to be fair
- Other issues with justice (i.e. fairness) seems of little importance in many situations
- Counter-intuitive view of supererogatory action
- Demand for perfect impartiality counter-intuitive (demands seem inappropriate or even impossible)

热门
对……的异议
功利主义

- 这是一种猪的学说。
- 缺乏计算效用所需的时间
- 对人类而言要求过高 y
- 动机性／心理性享乐主义似乎存在谬误
- 规范性享乐主义似乎存在谬误
- 欲望满足论的福祉观（密尔的证明）并不成立
- 效用加总往往导致反直觉的结果
- 强版本的消极责任学说存在谬误
- 各种琐碎行为被错误地归类为道德上错误的行为
- 信守承诺似乎并无多大重要性
- 惩罚未必需要公正
- 在许多情境下，正义（即公平性）方面的其他问题似乎并不重要
- 对超义务行为的反直觉看法
- 要求绝对的不偏不倚显得反直觉（此类要求看似不合时宜，甚至根本无法实现）