Contents

[Introduction: Why are people utilitarians? 3](#_Toc71712473)

[Sophisms 3](#_Toc71712474)

[Harm and necessity 3](#_Toc71712475)

[The internal properties of utilitarian’s 5](#_Toc71712476)

[Measurability 5](#_Toc71712477)

[Basic states 5](#_Toc71712478)

[Nietzsche 8](#_Toc71712479)

[Consent 11](#_Toc71712480)

[Two examples where the Utilitarian does not value consent 11](#_Toc71712481)

[Colonoscopy 12](#_Toc71712482)

[Weight and measurement 13](#_Toc71712483)

[Additivity 14](#_Toc71712484)

[Can we get intrapersonal? 14](#_Toc71712485)

[Interpersonal Harsayani 14](#_Toc71712486)

[Diminishing marginal utility 16](#_Toc71712487)

[Unicity 17](#_Toc71712488)

[Everything is obligatory 17](#_Toc71712489)

[Notes on the doctrine of double effect 18](#_Toc71712490)

[Classical Utilitarianism versus existence 18](#_Toc71712491)

[Preference utilitarianism versus existence 19](#_Toc71712492)

[Character 20](#_Toc71712493)

[Feelings that can count but do not change the outcome 20](#_Toc71712494)

[Feelings that should NOT count? 21](#_Toc71712495)

[Truth 22](#_Toc71712496)

[Utilitarian’s and lying 22](#_Toc71712497)

[Utilitarian’s and the gestapo 23](#_Toc71712498)

[Instrumentality of truth 24](#_Toc71712499)

[Justice 24](#_Toc71712500)

[Utilitarians judge 24](#_Toc71712501)

[The entailments of utilitarian’s 26](#_Toc71712502)

[Eugenics 26](#_Toc71712503)

[Value hierarchy 27](#_Toc71712504)

[Speciesm is bad 28](#_Toc71712505)

[Duties 29](#_Toc71712506)

[Utilitarians and the urbans 30](#_Toc71712507)

[Blue pilled 31](#_Toc71712508)

[Defining Human 32](#_Toc71712509)

[Humanity is second 33](#_Toc71712510)

[Particularism and Singer 34](#_Toc71712511)

# Introduction: Why are people utilitarians?

# Sophisms

### Harm and necessity

I want to evaluate a common argument I hear against meat eating:

**Argument**:

1)Causing unnecessary harm is bad

2)Eating animals causes unnecessary harm

→Therefore, eating animals is bad

Though the argument appears straightforward, it is nevertheless worth clarifying some terminology. The term “bad” is used in a strong sense of “we should not do what is bad”. In other words, bad is carries with it a kind of moral force, implying that if somebody accepts that an action has this property they should be compelled not to take that specific action.

Any one act can have numerous positive and/or negative effects and the argument can still maintain meaning. The simplest way these effects can be aggregated is through additivity, i.e. if one has reason to believe these effects can be added up then one can claim the action is bad in total.

An example of this is when the act of helping causes harm that is unnecessary for the consequence one aims to achieve with their action. For instance, perhaps I push you out of the way so that you won’t get run over by an incoming car, but as a consequence of my action you lose your hat. It is true that you losing your hat not necessary for being saved from an incoming car. Nevertheless, we can judge a posteriori that it was a necessary consequence of the action taken.

An action does not always have to have the same desirability. There may be cases where an action is good and cases where that same action is bad. As such the above argument can be revised to apply only to certain situations. For example, if the action is to “give morphine”, it would depend on whether the person in question is in pain or not.

Harm is an ideal term to use because of its generality. The argument would appear to work if the words “pain” or “suffering” were used. However, the use of alternative words might exclude the concept of “killing”. Using those words would then make the argument more open to objections by way of empirical methods. For instance, one could just point to some painless way of killing and the argument would instantly fail. The harm formulation can survive such an empirical point. On the other hand, if one attacks a pain formulation that would probably also apply to the harm formulation.

This generalization also corresponds to what most vegetarians actually hold as their belief. If there was a very ethical farmer that practiced the painless killing of an animal, it is unclear and rather doubtful whether this would have an impact of the perspective held by many vegetarians enough to change their beliefs and proceed with eating that animal.

The problem of “harm” is that it may be general enough to encompass non-conscious agents, such as killing a plant or tree. The argument as presented refers only to animals but may be opposed by asking: “why only animals and not plants?”. In this case, the vegetarian may return to the previous standards of “suffering” or “harm”. Alternatively, they could commit to “mammalianism”, that is, arbitrarily discriminate between organic beings based on whether they are mammals or not.

However, I suspect the most likely position they will take is that it is not a matter of category but a matter of degree. That is, they will agree that harming plants is bad, but not bad enough considering the benefits of the action. That is, one may think that the value of plants is high, but the value of human living is higher. I believe this kind of position automatically locks into additivity.

**The Problem**

The most obvious problem with the argument is the notion of necessity. A circular definition of the term is: X is necessary if the presence of X is required for a certain other thing to occur. It makes little linguistic sense to talk of necessity without a cause, necessity is a constraint and there must be some objective for the constraint to work on.

For instance, if I want to make a cake it is necessary that I use the ingredients necessary to make the cake. The sentence “flour is necessary to make the flour cake” makes sense. The sentence “flour is necessary” does not make sense in itself. Thereby, it becomes clear that vegetarians are assuming that there is some goal, in this example the cake, which can be achieved through a variety of means.

What is the “cake” of the harm done to animals? Suppose an agent is trying to get the best “taste” possible, the so called “omega taste”. If the omega taste does not require eating animals, then the argument works. This would be equivalent to saying: “don’t eat animals because there are better tastes out there”. If on the other hand the omega taste must include animal flesh, then the argument instantly fails. That is, if I am trying to have the best taste I can, then it IS necessary that I eat animals.

I suspect that the herbivores then have a rather different perspective. They are instead redirecting us to look for another “cake”. A reflexive reaction to this might be “who are they to tell us what our goals should be?”. Perhaps they know better than us, either because they have reasoned better or maybe they have information we do not.

If it is true that they have reasoned better than us, then perhaps this reasoning can also be shared with us, and upon engaging in this reasoning, we can be persuaded to also change our goals. If on the other hand they have better experience than us, then they need only try and help us experience those same things.

Perhaps more importantly, necessity is dynamic, not static. In a dynamic world, there is a starting position and there are costs associated with transitioning. For instance, it may be that what I want is to live in a community where people look out for each other, it may appear that if I get rid of say bullfighting, we can find other ways to bond, but in practice, when bullfighting was abandoned, the community came apart. Saying something is not necessary misses the point about what paths one can safely transition to. It may be true that other paths can achieve the same function but they are not available from the current state, or at the very least not available without significant damage.

To summarize, the argument fails because it in fact totally ignores the function of the behavior.

# The internal properties of utilitarian’s

## Measurability

### Basic states

Utilitarians are fundamentally interested in states of affairs, not actions. The basic state of affairs they are interested in is the pleasure and pain tradeoff. An action is itself never morally good or bad, it is so in virtue of the quantity and quality of the pains and pleasures it brings.

The problem is that the states in which utilitarianism is interested, are fundamentally unmeasurable. The basic goods which it wants to structure life around are merely inferred through manifestations like yelling or orgasmic appreciation.

Pain is arguably easier to measure, so in a society that is profoundly utilitarian, you would expect people to maximize their expression of pain to maximize attention on themselves. But even with pain, the relationship between pain felt and the manifestation of pain is not simple. In Daniel Kahneman’s famous experiment, people were put through 10 minutes of hardcore colonoscopy and some other people were put through 10 minutes of hardcore colonoscopy AND another 5 minutes of softcore colonoscopy. The latter group reported a less painful experience. How does a utilitarian minimize pain in this scenario?

Pleasure is even more difficult. As an example, it seems men are naturally more stoic, are we to infer that they feel less pleasure? It seems that fundamentally, two people in identical positions physical positions, one of them could be feeling immense pleasure and another a milder or even moderate pain.

Yet despite these issues, utilitarianism insists that the basic goods are these non measurable states. This is a recipe for conflict and widespread disagreement, it is perhaps no coincidence that the century is rife with psychologists, psychoanalysis, and neuroscience. These are, to the utilitarian, the most important things that society can put its attention to. The person who has the best measure of these states is the person we should be listening to the most. No wonder this century has seen such a massive profileferation of psychologism.

Yet the focus on these states of affairs inevitably leads to conflict as people can disagree about which state of affair is more desirable. For instance, a teenager may say that the utility of visiting your grandma on the weekend is lower than the utility of going skateboarding. If his parents disagree, is there a way to resolve this moral dispute?

The teen may say that going skateboarding will allow him to make friends and maybe a girlfriend so more people will benefit overall. The parents may say that the grandma does not receive many visitors and hence a single visitor will have a larger impact than the marginal impact of him joining a group of friends. Both of these seem like reasonable arguments, in a non-utilitarian framework they may not even be moral disagreements but in the utilitarian framework, this IS a moral disagreement.

To simplify, ignore the brain state of the teenager. What is being compared, is the brain state of the grandmother from getting a visit from her grandson, to the brain states of the teenager’s friends. Is there a way to find out which brain state will produce more utility? It doesn’t seem like there is. It seems particularly odd to me that a utilitarian could believe that psychologists and the like will one day resolve this dispute, but I certainly do not.

Is there a way to resolve this dispute for a utilitarian? It does not really seem to be so. This is the basic problem of measurement. Utilitarianism posits an invisible non-measurable state as the optimand. A utilitarian can try to claim that the psychologist will one day be able to measure some of the states in question. But is there a reason to believe that what we CAN measure is more significant than what we cannot?

Even if we can get weak measurements there seems to be a tendency of experts to favor the more robust measurements. The reason people are always obsessing over income and happiness is not because it predicts happiness better than other measures, but because other measures have a lot of ways of being measured. The expert has a tendency to dismiss qualitative data, merely because it is harder to measure.

Let us take an example. Suppose that Carna has a butcher, Butch, who is friendly. When Carna visit him she ends up talking to him for an hour before buying beef for 15 euros a kilo. There is also a cheap, impersonal butcher, Cheapy that is selling the beef for 10 euros a kilo but Carna does not visit Cheapy. Perhaps we can infer that the value that Butchy gives a value higher than the 5 euro difference, but we really have no idea. Carna could be valuing Butch’s service at 100 or 200 or could hate Cheapy because he cheated on her.

Worse yet, if Carna was a utilitarian, she would try to compartmentalize value and optimize her value. A simple attempt to optimize by Carna could look like this: ‘Why don’t I buy the cheaper steak, and go talk to Butch anyway?’. Needless to say, thinking too much in this profit and loss framework can be a recipe for value destruction, in the end, Carna could cause Butch to close up shop.

The economist will then come in and measure things. The economist will reason that before she was paying 15 euros for a steak, and now she is paying 10. That must mean that her cost of living has decreased, which means she has more disposable income, which means she has a higher standard of living. Yet, we know that this is not necessarily so, her conversations with Butchy have been eviscerated. Measured value has gone up whilst real value has gone down.

This seems to be exactly the problem with contemporary society. Intellectuals cannot measure community ties and family bonding or integration so they focus on material consumption, claiming that this is sufficient to show that life is better today. Yet, this is absurd on its face.

A utilitarian may claim that these disagreements are exactly what the public square is about, ‘debate is the instrument of measure, we can resolve our disagreements by arguing’. But this opens the door to another objection, measurement itself may be costly. Suppose we try to sit down and try to have a conversation about what has a higher value, we do this often. Perhaps this sometimes works and we resolve it, and at other times it does not. But by engaging in this conversation we have stopped other activities from taking place. Who is to say that the utility of having these conversations is higher than those other things we would have done? In fact, this leaves open the possibility that being a utilitarian may in fact NOT maximize utility because it would advocate that we spent more time arguing than maximizing utility, even if we were doing so by accident.

In conclusion, utilitarianism posits that the ethically important thing, exists, is partially measurable and that which IS measurable is more important than what isn’t without basis. Utilitarianism is the disease of the enlightenment, institutions failed to compartmentalize the scientific method to the domain in which it is applicable and brought it into the public square, resulting in rather absurd inferences and collapse of moral intuition.

**Edit**: Some dialogue to go with this. See [here](https://twitter.com/diomavro/status/1384800640233246720?s=20)for animation

Sophist: Look… it’s all so simple. We should maximize the total pleasures minus pains.

Socrates: Let us say we agree. What next? How do we know what has more pleasure or pain?

Sophist: What do you mean? Isn’t it obvious when somebody is in pain?

Socrates: I don’t know, people can act like they are in pain or can hide their pain. Ever heard of Goodhart’s law? If you are targeting their pain, you are giving them a reason to act like they are in pain.

Sophist: Eh… but we can measure pain objectively, I mean… not all pain… but neuroscience and psychology are able to make scientific measurements of some of them.

Socrates: Wait… are you saying we can’t measure all pains?

Sophist: Well yes, it seems like some of them are more complicated to measure.

Socrates: And how do you know that those which we cannot measure are less important than those which we can measure?

Sophist: Well look… we have to make assumptions somehow!

Socrates: You are changing the optimand! Instead of saying we should focus on pleasures and pains, you now want to focus on MEASURED pains and pleasures, even if it means LOWER pleasures and pains OVERALL. Just because people have more iphones, doesn’t mean they are happier… how can you weigh material goods against depression and suicide.

Sophist: No… you don’t understand the second law of thermodynamic… it’s science! It can solve everything!

### Nietzsche

Most Christians acknowledge Nietzche as one of their most astute critics. Yet, for his time, he spent a surprising amount of space critiquing Utilitarians. Indeed, from his text, he seems to imply that Utilitarianism adopts all the worse aspects of Christianity but then also piles on top of it, an almost caricatural understanding of pleasure.

Nietzche contends that the Christian and the Utilitarian make the same mistake: they universalize beyond their own place in society. The Utilitarian naturally adopts the morals he does because he is an aristocrat who wants to impose something on the classes which he dominates, yet he also desires that those classes themselves adopt the utilitarian doctrine, which will make the lower classes more submissive to the imposition of the aristocrats. The Christian adopts the morality of a good slave, yet desires his master to also adopt that same ethic.

The difference between the Christian morality and Utilitarian morality is that the Christians seek to suppress some sorts of pleasure whilst Utilitarians encourage it. Yet he sees the set of pleasures that Utilitarianism seeks to allow as pleasures of the lowest human, the “herd animal”, the “boring” and “mediocre” enjoyment of people who have yet to awaken from the “soporific” spell of slave morality (Beyond Good and Evil 228).

Nietzche thinks that the function of morality is to help the herd. Utilitarianism fails at this in two ways: one we can call the overgerenalization of equality, and the other the Hedonic Paradox.

The first is that it wants to apply morality BEYOND the herd, to treat non-equals equally. For example by demanding that all utilities be considered, not just the utilities of the herd, it demands an extension that will simply ruin the function of morality. I think in economics the parallel would be if, say, Microsoft valued the profits of Apple as much as it does its own. This kind of perspective destroys the function of competition. A requirement of impartiality requires that we assume equality even among non-equals, clearly an inheritance of Christianity. In the Will to power he says “It is the instinct of the herd that finds its formula in this rule, one is equal, one takes oneself for equal”, and then identifies this same principle working in Bentham under the dictum,”Everybody counts for one, and nobody for more than one.”

The reason this is the mentality of the herd is due to the benefits of cooperation in a group. In other words, it is a rule of thumb about how to behave within a group to avoid failing to cooperate. Yet as a general rule it is simply an overgeneralization, “John Stuart Mill believes in it” as the basis of morality, but that he fails to grasp its prudential origin (Will to power 925). Pleasure and pain play the same role, they are rules of thumb that a herd has used to survive effectively. That is, pleasure and pain are INSTRUMENTALLY useful to the herd as tools for survival and growth. In that sense, prioritizing pleasure and pain is like an army prioritizing the sharpness of their swords instead of prioritizing victory. Part of this critique is the idea of diminishing marginal utility. Since Nietzche thinks of morality functionally, he sees no reason to suppose that one person will do less to enable the good life than two. Indeed some people think that his optimand is to create the best element possible, create the best man it is possible to make! Happiness is simply useful to the cultivation of human excellence!

The second failure of Utilitarianism is the psychological effect, also known as the Hedonic Paradox. Utilitarianism makes two propositions: that we should focus on aggregate happiness AND that focusing on aggregate happiness will increase aggregate happiness. Yet, by avoiding pain, suffering, and inequality, the Utilitarian will harm the survival function of morality. He thinks that one ought to emphasize suffering and intellectual struggle for advancement of the individuals in the groups, and hence for the group as a whole. So Utilitarianism has an internal struggle it can never overcome: its adoption will lead to its own negation. It sets a goal for itself, and then gives a method which will fail to achieve that goal.

This critique seems to apply to Bentham but not to Mill, who distinguished higher and lower pleasures, and who admits the non-linear method of increasing the aggregate. Sidwick, another well known utilitarian also seems to embrace the notion that reasoning in Utilitarianism may not be optimal: ‘Utilitarian may reasonably desire, on utilitarian principles, that some of his conclusions should be rejected by mankind generally [if their rejection will ultimately lead to greater aggregate happiness].” We can summarize this point as follows: Utilitarianism suffers from Goodharts law. If it becomes a target, it won’t be a good measure. Let those OUTSIDE the group measure our happiness, but if we ourselves maximize our happiness, it won’t work.

So this second critique undermines Utilitarianism as a decision making procedure, NOT as a standard of value. For instance, it is good if we have aggregate happiness but it could be bad to have a utilitarian population. The utilitarian population which attempts to frantically measure and optimize pains and pleasures does not seem to be effective. Part of the psychological critique is that by adopting a clearly stated morality, there is a tendency to assume that the rules we need to survive are timeless and universal, which will reduce our adaptability.

##### RELEVANT CITATIONS ON THE FIRST TWO PARAGRAPHS:

Passages Notable quotes from Beyond Good and Evil

*“slave morality is essentially a morality of utility”*

*Nietzche characterizes the genesis of Utilitarianism as the aristocratic man who contrasts himself with “the cowardly, the timid, the petty” or “those who think only of narrow utility”*

*And adopts the Christian “pity, the kind and helping hand, the warm heart, patience, industriousness, humility, friendliness come into honor? for these are the most useful qualities [for the slave]”*

**{BGE 260) (Beyond Good and Evil 260)**

*“English happiness,” namely, “comfort and fashion”*

**{BGE 260) (Beyond Good and Evil 228)**

##### RELEVANT CITATIONS ON THE REST:

*If you experience suffering and displeasure as evil, worthy of annihilation and as a defect of existence, then it is clear that besides your religion of pity you also harbor another religion in your heart that is perhaps the mother of the religion of pity: the religion of comfortableness.*

*(of Bentham) How little you know of human happiness, you comfortable and benevolent people, for happiness and unhappiness are sisters and even twins that either grow up together or, as in your case, remain small together.*

*It never occurs to them that the path to one’s own heaven always leads through the voluptuousness of one’s own hell.*

**Gay Science 338**

*This religion? Or, more specifically, morality of comfort thwarts its own goals by attempting to eliminate all suffering.*

**Beyond Good and Evil 44**

For some reason I can’t download the PDF but inspiration of this post comes from a paper called: Nietzsche’s Critique of Utilitarianism by Jonny Anomaly

### Consent

The Utilitarian does find value in consent, but the value is not intrinsic but instrumental, it comes either from game theory or from information. The game theoretic view can be illustrated as follows: If a woman knows that people will only have sex with her if she consents to it, then she will walk on the street without worry. The Utilitarian believes that there is more total pleasure if she has the ease of mind to walk on the street. That is, requiring consent as rule, allows people to cooperate on a better equilibrium.

The second way Utilitarians value consent has to do with imperfect information, what economists call, revealed preference. That is, we don’t know if somebody prefers A or B, unless they have the choice between A or B and choose A. For instance, if they can choose vanilla or chocolate, and choose chocolate then we can say they prefer chocolate.

But the economist framework goes further, it says that they are indifferent between receiving chocolate AND choosing whether to have chocolate or vanilla. That is, there is no value on the agent having a choice as such. So if we can predict your choice, there is no point in giving you your choice! (economists call this the [independence of irrelevant alternatives](https://en.wikipedia.org/wiki/Independence_of_irrelevant_alternatives))

Can we try to adapt the framework? What if A = ‘having the choice between chocolate and vanilla’, B = ‘being given chocolate’ and in this case the agent prefers A. While this formalization might be interesting, it is no longer Utilitarianism because by allowing pleasures to NOT be about consequences but about the arrival to the consequences, we have kind of abandoned the whole framework.

### Two examples where the Utilitarian does not value consent

There are many examples where the Utilitarian does not value consent but let us focus on two: 1) He or she ignores consent and saves somebody who does NOT want to be saved, under the condition that saving them would add to the total pleasure. 2) He or she ignores consent if somebody does NOT want to be killed, but their being alive decreases total pleasure.

Imagine a ship at sea that is sinking. Everyone is getting on their lifeboats but the captain refuses to do so. He believes that he ought to go down with the ship since it is his responsibility. This is not based on some utilitarian way of reasoning, indeed, if he were to live, he would probably have a fine life and even bring happiness to those around him. This is simply a question of duty. Now, it so happens that you have a tranquilizer gun and you can tranquilize him and put him on a boat. Should you do so?

The Utilitarian doesn’t seem like he would value consent in this case. I think a non-utilitarian may come down on either side of this. A more straightfoward example of this case is the girl from the short story ‘[The ones who walk away from omelas’](https://www.goodreads.com/book/show/92625.The_Ones_Who_Walk_Away_from_Omelas)

Let us take the example of somebody who does not want to be killed but the utilitarian is morally obliged to kill them. It could be a [young](https://www.youtube.com/watch?v=7-w6c-ybwXk) [person](https://www.bbc.com/news/stories-45117163)but let us stick to the more common example: An older person truthfully claims that they are suffering a lot more than they are feeling pleasure. They may say that they don’t enjoy the taste of food anymore, they are too weak to walk, they cannot properly see their grandkids anymore, and they think they make the grandkids miserable when they visit. However, the catch is, this older person is ethically against suicide. Should you unilaterally kill them?

Why does it have to be unilateral? Because if it wasn’t unilateral, if they consented to be killed, then it would be suicide, you would only be an instrument to their suicide and that goes against their beliefs. So it seems like a Utilitarian would in fact be morally obligated to go around killing these sorts of people. To avoid shocked relatives, it’s best to make it look natural.

For a little bit of contrast, I would say that the traditional imperative to ‘respect’ your elders, is not in fact asking you to calculate their pleasures and pains and act based on the sum but asking you to respect their WILL, which has no value in a utilitarian framework.

### Colonoscopy

Daniel Kahneman has a famous colonoscopy experiment which he also describes in his book, thinking fast and slow. In the experiment, he splits patients into two groups and gave each group a different treatment. The first treatment is a colonoscopy which is intense for some period of time, X. The second treatment is a colonoscopy which consists of the first colonoscopy, plus another 10 minutes of less painful colonoscopy. The patients are then asked to rate the experience. It turns out that those who suffered more in total, preferred the experience to those who suffered less. Kahneman summarizes his results by saying that people do an average of the ending pain and the maximum pain when thinking of their experience.

This puts utilitarians in an odd position. Which treatment should be preferred? It seems like one group had a lower experienced pain and a higher remembered pain, whilst the other group had a lower experienced pain but a higher remembered pain. So now the utilitarian must make a decision, which is it? Presumably there will be some weighting function between these two pains, minimizing pain is simply not enough to prescribe.

It is also interesting to wonder if this effect would remain if it were explained to the patients what the options were. Would the effect disappear? If so, we might have another incentive for lying.

### Weight and measurement

Suppose that there are 3 types of objects and somebody can only consume 5 objects in their lifetime. Also assume that everything is consumed instantly at one point in time. There is only one person in existence, and he will live for one second, where he will consume all 5 items. We want to figure out what the 5 items this person will consume are.

The brute force way of getting this is to ask for his preferences over all possible combinations. So (5,0,0), (4,0,1), (4,1,0), (3,2,0), (3,0,2), (3,1,1), (2,3,0), (2,0,3)… or we can use [Feller’s](https://en.wikipedia.org/wiki/Stars_and_bars_(combinatorics))method to calculate it and get (5+3-1)!/(3-1)!=49. So we need to know this persons preferences over 49 combinations. But this is too much work, maybe there is a simpler way of getting preferences over this set?

To develop our intuition, suppose we were simply trying to find out the heaviest bundle. Then, we would only need to know the weight of A, the weight of B, or the weight C. In fact, we don’t even need to know the exact weight, we only need to know that A is heavier than B, and that B is heavier than C! Once we know this, we can also infer that two A’s are heavier than two B’s, and that one A and one B are heavier than two B’s! In fact, with just the order A>B>C, we are already able to say what the heaviest basket is, five A’s.

So back to the initial question, what if we were trying to get the basket with the highest utility? If we are given the preferences A>B>C, can we infer anything about the best basket? Well no… it may be that B and C are complementary or it may be that utility is increasing in the number of C’s or it may be that A and every other good are substituable. We can’t conclude anything about the best basket of goods, we can’t even conclude if A should be in the final basket!

What if this person was somehow able to not give us an ordinal ranking but was able to precisely measure their units of pleasure from each item? They tell us, A gives me 10 pleasure, B gives me 5 pleasure and C gives me 1 pleasure. Can we now make inferences? Well no! The problem remains, we don’t know how much pleasure the extra units and the combinations of these goods will bring.

In other words, when the thing we are trying to maximize is utility, we cannot make a measure that reduces the amount of measurement we have to do. With weights, we don’t even need to have a cardinal measure of the objects. With utility, even if we had cardinal measurements we would fail to get the best basket.

For more I recommend reading Alchian’s [paper](http://www3.uah.es/econ/MicroDoct/Alchian-Utility%20Measurement_1953.pdf), ‘The meaning of utility measurement’

## Additivity

### Can we get intrapersonal?

A Neumann Morgenstern (Hungarians think it is pretentious that he added a Von in front of his name when he got famous) utility function is a somewhat odd creature. The input is ordinal, you have somebody’s complete, transitive, continuous and independent preferences and then you can construct the utility function. So we have a series of objects to be compared, A, B, and C.

Let’s give a brief overview of the axioms. Complete means that you can compare all of them. That is, you can say which one is weakly preferred to each other one. If A is weakly preferred to B, and B is weakly preferred to A, then you are indifferent. Transitive means that if you weakly prefer A to B, and you weakly prefer B to C, then you must weakly prefer A to C. Continuous means that if you have preferences A>B>C, then there must exist a p, such that you are indifferent between receiving B or the lottery pA+(1-p)C. Independent means that if you prefer A to B, then you should also prefer the lottery (0.5A+0.5C)>(0.5B+0.5C).

If these axioms are present then you can represent the person’s preferences using a utility function. The utility function can assign numbers to each good and perfectly capture the ordinal preferences. So A=10, B=5, C=2. These can be used as a substitute with the ordinal preferences of the person in question.

However, even though these are numbers and the utility function itself is cardinal, the inputs are STILL ordinal. This means that it is STRICTLY nonsense to assume that A is preferred five times more to C. The representation is cardinal but we have NO cardinal information. In fact, in these utility functions, it is mathematically equivalent to always give the highest preference a 1, and the lowest preference a 0. But again, this does not mean we can claim that A is infinitely many times preferred to C. It is just a number chosen for convenience.

The theory behind the constructions of utility functions are about how we can use utility functions to represent ordinal preferences, and they do not claim that the utility functions capture all of the information. For the constructed utility functions to actually capture people’s preferences in a way that is meaningful, it would have to be that there exists only ONE utility function that can capture somebody’s preferences. To repeat myself, if you can use two utility functions to capture somebody’s preferences, this means that there is no MEANING behind the cardinality.

### Interpersonal Harsayani

One of the most famous defenders of utilitarianism was the nobel Prize winner in economics, Harsanyi(Hungarian btw). He had claimed that his theorems were the heart of utilitarianism. Yet the Neuman Morgensten axioms they used to construct their theory, were by construct, ordinal, which was odd because utilitarianism as an ethical theory requires cardinality. This was, at the time, seen as absurd by people such as Alchian, Ellsberg, Friedman, Savage, etc that these utility functions should NOT be confused with the classical interpretation of utility that the neoclassicals used. Ultimately the critics were right, you cannot construct a measure of utility without a reliable way of measuring cardinality (I will remind you that the level of happiness research is asking people how happy they are on scale of 1-10).

Unfortunately, because of the sales pitch of Neuman and Morgensten(‘that allows them to put the measurement of utility on as firm a foundation as the measurement of, say, heat ‘) and because Harsanyi didn’t formalize his ideas properly, it took a while to figure out what was going on. About 30 years elapsed since Harsanyi’s radical claims that these utility functions did have welfare implications and the time we were finally able to say that the assertions are false. Along the way, a lot of economists got confused about what was possible, especially those who were simply following from the side lines, as at the time, many economists were not trained in real analysis to be able to read this literature themselves.

So what did everyone agree that the axiomatization achieved? The purpose of the theory was to link preferences to affine functions, the functions were chosen because they are analytically tractable. In other words, the construction of such a utility function is meant to describe the behavior of agents. It was not meant to have any normative content.

Harsanyi’s contribution is to try to extend the theory to do more than describe. He wanted to show that if you aggregated these utility functions, they would, in fact, represent utilitarianism. He did this using two theorems, the impartial observer theorem, and the aggregation theorem.

In the impartial observer theorem, Harsanyi takes a very Rawlsian approach (it is before Rawls). He imagines that if everyone has an affine utility function, and some impartial observer imagines himself having an equal probability of being any one of those agents, then he would simply maximize the average utility function and that’s how you can construct a social welfare function.

His aggregation theorem, is a bit distinct. He proves that if the individual utility functions that people have satisfy the NM or equivalent axioms, then the impartial observers utility function also satisfy those axioms. This isn’t always possible, it requires proof, often aggregation can lead to different properties emerging. Part of why Harsanyi’s work is important is because he found a way to aggregate where you don’t lose the initial properties.

The problem of course, was the starting point of Harsanyi, the NM utility function. The problem is that it is not unique, in fact, arrow demonstrated a bunch of other kinds of utility functions that had as desirable properties. Ultimately, these functions were chosen not because they are more likely to be true, but because they simply easier to compute. But this means that the function you have chosen, is in fact, arbitrary.

Worse yet, apart from there being many utility functions, even if you started out with just the NM utility function, there would STILL be numerous social welfare functions. An additional axiom is needed to make a unique one, the axiom is called ‘Independent Prospects’.

In essence, the confusion stems from the apparent cardinality of the utility function. In fact, just because you transformed an order into a cardinal representation does not mean you have cardinal information. It is very difficult to actually make a utility measure which is of [greater use](https://diomavro.wordpress.com/2021/05/10/measuring-utility/).

The idea to make these functions is to say that, person A’s favorite state is 1 and his least favorite is 0. The same thing can be said about person B. We can say that A prefers 0.5 to 0.4, (intrapersonal comparison) but it is nonsense to say that A having 0.5 and B 0.3 is worse or better than A having 0.3 and B having 0.5. This is technically impossible, indeed, for person A. This is especially strange because utilitarians DO claim there are higher pleasure and lower pleasures, that is, reading a good book is not equivalent to scratching your head, yet if A is a dog and B is a human, and 0.5 for the human is reading a book, we are required to give equal weight to the dog scratching there head as the human reading the book. If we DON’t do the 0-1 normalization, then it’s open season, it’s even more impossible to compare utilities!

Ref: Weymark, John A. “Measurement theory and the foundations of utilitarianism.” Social Choice and Welfare 25.2 (2005): 527-555. It is [here](https://link.springer.com/content/pdf/10.1007/s00355-005-0017-7.pdf)

### Diminishing marginal utility

If you take a billion euros from Bezos and distribute it to every American, they should all see around 3 dollars. More importantly, according to a utilitarian, total utility will increase. The reason it will increase is due to diminishing marginal utility, that is, a rich person gets less from the marginal dollar than a poor person.

But if you believe this is true then another example is also true. Suppose now that we can televise Bezos’s hand getting chopped off. Bezos is not particularly troubled by this, he doesn’t really need his hand, he thinks he can replace it with a robotic one, moreover, even if the robotic one fails, he has enough money that he can hire a right-hand man.

Now it turns out, that half of America will be greatly entertained by seeing Bezo’s hand get chopped off. In fact, they are ready to pitch in 20 dollars to see it. The other half won’t be watching so it won’t bother them, but they will receive 5 dollars from the other half who do wish to see it. This redistribution is the only cost that will actually be incurred, in other words, even though half of Americans are ready to pay 20 dollars to see Bezos hand get chopped off, they will only pay 5 to appease the other half. In other words, everybody is better off in this equilibrium.

It seems to me that if somebody thinks that taking Bezos money is welfare increasing, they are committed to saying that chopping Bezos hand is even better. In fact, this is just one example, I will leave it to the reader to imagine how we can justify brining back gladiator games.

If you want to read more about this, see Sen’s article, ‘Utilitarianism and Welflarism’

## Unicity

### Everything is obligatory

Other ethical systems may 1) give you rules on what not to do, or 2) may outline criteria after which an action becomes morally permissible, in contrast, utilitarians claim that everything they say is a moral obligation. If something maximizes utility, it is your obligation to do that, there is no room for disagreement or tolerance. If you are utilitarian, and somebody is not maximizing utility, they are being unethical. This means that it has no room for pluralism in ethical values, there is no room for character at all. In fact, the best moral agent to a utilitarian is a robot, any kind of subjectivity in calculating the right action can only lead us astray. Whilst classical ethics, in the form of virtue ethics, actually allow one to look within for the answer, utilitarianism focuses on a vague cardinality of future pleasures and pains, which are best quantified, and hence take the human out of the equation.

Since utilitarians are likely to acknowledge that you can do more to maximize total utility by focusing more on the utility of others than your own, they also in fact prescribe that everybody do the same thing. That is, if volunteering at charity X does more to help total utility, then everyone should volunteer to work at that one charity. Their only mechanism for generating diversity in actions is diversity in competencies, they may say, if you are a good investment banker, focus on doing that and then give your money to the most effective charity. The ideal utilitarian simply earns a lot of money, even by wronging people, and then simply gives that money to the most effective charity(since that charity specializes in maximizing utility).

So by definition, if Bob meets Anna, Anna is a utilitarian, and Bob agrees with Anna that his actions are not maximizing utility, this implies Anna thinks Bob is unethical. Of course in practice, utilitarians will try to hide this aspect of their view by simply saying that you are too ignorant so you are only acting under incomplete information. In sum, a utilitarian thinks non-utilitarians are either unethical, or idiots.

I do not think this requires any more elaboration but maybe for constrast let us simply compare this to a rule like the doctrine of double effect. The modern elaboration of the doctrine was made by Phillipa Foot, when she introduced the Trolley problem. Her formulation of the doctrine is still being debated but what I want to focus on here is that she has her four criteria, which I include below, and IF those criteria are met, the act becomes morally permissible. This means Foot would never say you have a moral obligation to kill somebody, simply that killing somebody can become permissible. This means that her view can be scaled in a society such that people who hold various answers can co-exist(in the sense of find each other ethical).

## Notes on the doctrine of double effect

Foot in her original trolley problem article outlined the docrtine of double effect. The doctrine comes from Aquinas Summa Theologia and it goes something like this. For an action to be morally permissible, four things must be satisfied.

1. The action must be good or permissible (nature of act)
2. The bad effect must not be the means by which the good effect is achieved (means/end)
3. The intention must be only for the good effect, the bad effect is simply a side effect.
4. The good effect must be at least as important as the bad effect

Point 3 says that the END that is intended is the good and not the bad. An important discussion she has betweem action(direct intention) and allowance(obliquely intended). Allowance can be further broken down into letting things as they are, versus removing an obstacle or commiting an ommission. If we are deciding who to save, we can choose the higher number and if we are deciding what not to allow, we may also choose the higher number but we cannot compare the duty to save with the duty to help.

### Classical Utilitarianism versus existence

If you are a classical utilitarian there is somewhere strange diminishing marginal utility can take you. Since a classical utilitarian worries about the overall utility, including of not yet existing beings, then you have to take into account the effect of bringing more people into existence, and this usually leads to a paradox.

Let us imagine that we have a community of 10 people, each with 10 goods. If a new person comes into the community, then some of the goods will have to be distributed to the new person. If utility is diminishing then the allocation that maximizes utility with 11 people and 100 goods is that everyone get 100/11 goods, that is, perfect equality.

But this logic always works, in fact, bringing in a new person into the community when there is diminishing marginal utility, ALWAYS increases total utility. Take a simple square root as the utility function, if we have 10 people with 10 goods each, total utility is sq(10)\*10=31.6. With 11 people everyone gets 9.1, so total utility is sq(9.1)\*11=33.2. With 100 people, everyone gets 1 so total utility is 100. With 200 people, everybody gets 0.5 units, so total utility 141… this can go on until there are close to infinite people and everybody gets something approaching 0.

This is very close to the ‘need’ discourse. Utilitarians want everybody to be at the brink of starvation, it is inherent in the way they reason. Of course, a simple way around this is to drop diminishing marginal utility, but this is usually the main assumption that led them to adopt utilitarianism in the first place, this is the assumption that causes inequality to be bad, which for most, is a MUST part of the framework.

### Preference utilitarianism versus existence

There is another form of utilitarianism than one I have been discussing so far which is dubbed preference utilitarianism. This kind of utilitarianism says simply that we ought to increase utility by satisficing the highest preferences possible from existing beings.

Because it takes into account only the utility of beings that already exist, this allows it to go around some of the criticisms of classical (hedonic utilitarianism). That is, we don’t consider the potential utility of a being that does not exist, even if that being would have an immense utility.

This is usually the kind of utilitarianism that is mostly used for justifying abortion or infanticide (the kind that Singer supports). Indeed, abortion seems to be the main reason why many adopt it. They start out by wanting to support abortion and then figure out that other frameworks don’t help them. For example, classic utilitarianism would say that if the baby brings in more utility than it costs, then it should be brought about. If you pair this up with diminishing marginal utility, you quickly get a result that says that abortion should be illegal for most households that are richer than some threshold.

The preference utilitarian on the other hand will not consider the wellbeing of entities that do not have preferences. Since a fetus does not have a preference, it should not be counted, but its parents do have a preference, so what action should be taken with regard to the fetus depends solely on the parents’ preferences.

There is a kind of tension here between babies and animals. Singer was well aware of it. I suspect this is the main reason he dropped preference utilitarianism and adopted the classical view. Singer used to reason that animals DO have preferences; we know this because we notice they avoid pain or act in ways to ensure a certain environment. But any kind of procedure he describes for concluding that animals have preferences can also be used to deduce that babies have preferences.

Preference utilitarians still do the interpersonal comparison, and that is how they get that disabled people’s interests should be weighted less than able-bodied people. Specifically, a preference utilitarian may weigh a blind person’s preferences less than a non-blind persons interests.

I won’t offer a specific critique today other than to say that preference utilitarianism may actually avoid getting wrong answers like its predecessor, if only it resisted interpersonal comparisons. The combination of using preference utilitarianism and no interpersonal comparison is something economists dub the Pareto criterion. This is a much more defensible moral criterion, but it doesn’t always give one answer and usually constrains the utilitarian much more, which is why it isn’t very popular. I would actually say that the Pareto criterion actually coincides with the saying ‘do no harm’, which is much more lindy than most of these modern constructs.

### Character

This is actually just an entailment of a previous post I made where I explain how in utilitarianism everything is [morally obligatory](https://diomavro.wordpress.com/2021/03/22/utilitarianism-everything-is-obligatory/).

## Feelings that can count but do not change the outcome

Here I just want to pull out two examples that Bernard Williams uses in his critique of utilitarianism.

*George, who has just taken his Ph.D. in chemistry, finds it extremely difficult to get a job. He is not very robust in health, which cuts down the number of jobs he might be able to do satisfactorily. His wife has to go out to work to keep them, which itself causes a great deal of strain, since they have small children and there are severe problems about looking after them. The results of this, especially on the children, are damaging. An older chemist, who knows about this situation, says that he can get George a decently paid job in a certain laboratory, which pursues research into chemical and biological warfare. George says that he cannot accept this, since he is opposed to chemical and biological warfare. The older man replies that he is not too keen on it himself, come to that, but after all George’s refusal is not going to make the job or the laboratory go away; what is more, he happens to know that if George refuses the job, it will certainly go to a contemporary of George’s who is not inhibited by any such scruples and is likely if appointed to push along the research with greater zeal than George would. Indeed, it is not merely concern for George and his family, but (to speak frankly and in confidence) some alarm about this other man’s excess of zeal, which has led the older man to offer to use his influence to get George the job…George’s wife, to whom he is deeply attached, has views (the details of which need not concern us) from which it follows that at least there is nothing particularly wrong with research into CBW. What should he do?*

**Bernard Williams, Critique of utilitarianism**

*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 them several armed men in uniform. A heavy man in a sweatstained khaki shirt turns out to be the captain in charge and, after a good deal of questioning of Jim which establishes that 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 protestors 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, then as a special mark of the occasion, the other Indians will be let off. Of course, if Jim refuses, then there is no special occasion, and  
Pedro here will do what he was about to do when Jim arrived, and kill them all. Jim, with some desperate recollection of schoolboy fiction, wonders whether if he got hold of a gun, he could hold the captain, Pedro and the rest of the soldiers to threat, but it is quite clear from the set-up that nothing of that kind is going to work: any attempt at that sort of thing will mean that 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?*

**Bernard Williams, Critique of utilitarianism**

The utilitarian answers both of these in the same way, and the answer is of course, morally oblitatory. Geroge should take the job because it maximizes utility, and Jim should shoot the indian because it maximizes utility. The utilitarian can make no sense of the idea of integrity.

The agents in the above situations may consider different ways of thinking about the above problems. They will clearly have conflicting instincts at work. One of the considerations may of course be the utilitarian consideration. If the agent considers the utilitarian point of view and concludes that other considerations are not important, then he will simply act on the utilitarian framework. Are feelings merely of utilitarian value? If ones feelings are simply to be weighted, then one can take actions against ones feelings, alienating themselves from their own feelings.

## Feelings that should NOT count?

*Suppose that there is in a certain society a racial minority. Considering merely the ordinary interests of the other citizens, as opposed to their sentiments, this minority does no particular harm; we may suppose that it does not confer any very great benefits either. Its presence is in those terms neutral or mildly beneficial. However, the other citizens have such prejudices that they find the sight of this group, even the knowledge of its presence, very disagreeable. Proposals are made for removing in some way this minority. If we assume various quite plausible things (as that rogrammes to change the majority sentiment are likely to be protracted and ineffective) then even if the removal would be unpleasant for the minority, a utilitarian calculation might well end up favouring this step, especially if the minority were a rather small minority and the majority were very severely prejudiced, that is to say, were made very severely uncomfortable by the presence of the minority.*

**Bernard Williams, Critique of utilitarianism**

Should the feelings of the majority towards the minority count? I suspect most people would say no, but I do agree that the example seems a bit strange, it doesn’t seem reasonable that people could be living together and there is BOTH resentment AND no negative externality. I would perhaps adjust the example to be about isolated communities but it would be strange, why would the majority be exposed to the minority they find disagreeable? I guess journalists?

In general the utilitarian in the above examples can worry about two effects. The psychological effect, which is part of utility and the precedent effect, which is that once something is done, it becomes easier for others to copy that action.

Interestingly, for a utilitarian to worry about the precedence effect, he has to assume people are silly. Since clearly the utilitarian will ONLY worry if the precedence IF the action that would maximize utility in THIS situation would in fact, not maximize utility in a DIFFERENT situation. Then that means, to worry about the effect of precedence, the utilitarian must assume that say Jim killed the Indian, then some people will overgeneralize and assume it is okay to kill in other situations where it does NOT maximize utility.

## Truth

### Utilitarian’s and lying

Utilitarian types tend to worship science, when probed as to why they worship science they will give an answer resembling ‘because it is the most reliable way to approach truth’. But of course, these are not truth maximizers, they are utilitarians… so even if it is the most reliable way to approach truth, that truth is only valued for the utility it brings.

Utilitarians are aware that their position implies that is just a means to an end. They react to this implication in one of two ways. The Sam Harri’s path, which is that truth is in alignment with [welfare](https://www.goodreads.com/book/show/18869177-lying). Or you can take the position that it is [sometimes okay to lie](https://noahpinion.substack.com/p/yes-experts-will-lie-to-you-sometimes) like Noah Smith claims. The second position is just biting the bullet, so let us focus on the Sam Harris path.

**Utility and truth are always in alignment so you should always be truthful.** This is a particularly heroic assumption, indeed, probably the clearest example of this is to imagine somebody on their deathbed asking for glimmers of hope. I doubt even utilitarians would deny these cases exist, but maybe they can just make a weird sub clause for people on their deathbeds. To avoid this clause, let us take a more relevant case: There is an unprecedented pandemic, if the doctors don’t get the masks while they deal with patients, two thousand more people will die, so you want to make sure that normal people don’t rush to buy masks to ensure the doctors can be supplied. It should be obvious that a utilitarian would lie! Heck, the nice thing about democratic systems is that ‘trust’ is usually person specific, If Fauci lies about it, he can be replaced with the next useful idiot who doesn’t have a record of lying and bam, credibility is reset. Utilitarians can try to argue that it isn’t about trust in a person but about trust in the ‘institution’ but I think that this critique is overblown, most people usually forgive the lies and hope the next person of influence isn’t going to also lie.

**Utility is furthered more on average by truth-telling.**This point of view relies on the agent not being able to distinguish between situations where truth helps utility and situations where it doesn’t. Indeed, this seems to be one of the advantages of science, science helps us find state variables where it is easier to predict the consequences of our actions. So if it is true that on average, you should tell the truth, then a corollary of this is that as our understanding of the situations improves, our lying should increase.

It is also interesting to understand what this does to the probability of being truthful as a function of influence. Take Harris’s main reason for telling the truth, that it allows agents to form a relationship (which increases welfare). Now it seems quite clear that when somebody of influence speaks, there are not proportionately as many relationships being formed. It would be rather absurd to claim that Mehmet the Conqueror formed a relationship with everyone who listened to him. Instead it seems clear that when Mehmet speaks, very few relationships are being formed but there is a unilateral effect on a lot of people. So the tradeoff for a person of influence, seems skewed towards lying.

So to sum it all up, Utilitarians are telling us they will lie more, the more their understanding increases and they will also lie more the more influential they become. As soon as they have a position of influence, they will in fact lie. This is but an implication of what it means to value truth for its effects on pain/pleasure.

### Utilitarian’s and the gestapo

Today we are trying to do a bit of a steelman of utilitarians. What kind of assumptions do we need to get them to avoid telling lies? The economic view of truth telling is really quite simple. It is all about incentive compatibility, somebody will only give you the truth if it suits their ends. This means that before deciding whether to tell someone the truth, you need to know what the other persons intentions are. If you find their intentions are in harmony with your own, or better yet with everyone elses, you will share information.

Let us illustrate with an example, suppose that there are **citizens** and the **Gestapo** have come to ask them if they are hiding any **Jews**. The citizens want to protect as many Jews as possible, which means their ends are not aligned with the Gestapo. An entailment of this situation is that citizens want the Gestapo to have the least amount of information possible. The purpose of the Gestapo asking questions is that they can search more strategically than randomly, so they want to minimize the number of houses they search but maximize the number of Jews they catch, in the ideal situation, they would only search the houses with Jews. Absent any information, they will simply randomly search some houses.

We need to use a notion of equilibrium here. The equilibrium will be simply that nobody wants to change their behavior. To fix ideas, let us say that 10% of citizens are hiding Jews. When asked about the whereabout of Jews, they can say one of three things: ‘**I know** where the jews are’; ‘**I don’t know** where the jews are’; ‘**I am not telling you** what I know’.

Clearly, the ideal situation for the Gestapo is that they have as much information as possible. The most simple way this occurs is if the 10% who know where the Jews are, **truthfully reveal that**, and the other 90% say **they don’t know**. However the Gestapo doesn’t really need them to say the truth in this way. If all those who DO know say ‘I am not telling you what I know’, and the rest reveal that they don’t know, then the Gestapo will simply know that those who are not being cooperative are the 10% whose houses they want to search. Similarly, if everyone claims they are hiding jews, then the Gestapo will not listen to them. This because a policy where you search all the houses would be too costly, so the Gestapo would rather ignore everyone and search randomly than actually search any house.

For the Gestapo to get no information it must be that both groups say the same thing. We could have that both groups say they know, or they both say they don’t know, this would of course require that at least one of the two groups lies. However, if both groups say ‘I am not telling you what I know’, this also gives no information to the Gestapo and also ensures that nobody lies. So there are three equilibria where the jews saved is maximized. In my view, a utilitarian would seem to be indifferent between the three equilibria, but someone from a more classical tradition would prefer the equilibrium where everybody refuses to cooperate.

Is there a case where the utilitarian might prefer the equilibrium where everyone refuses to cooperate? Of course, one could just say that lying is costly and hence that equilibrium is less good but this is a little too easy, like saying that telling the truth is good in itself, which would make you a non-utilitarian. Let us instead add a secondary optimand, for the citizens, suppose that we are also trying to maximize the cost of the Gestapo. The Gestapo’s costs can go up in two ways, searching more houses, and punishing citizens. Punishing citizens can only be done when citizens are not cooperative, or when it was discovered that they lied. The people the Gestapo punish, the more costly it is, and the Gestapo have to treat everyone equally under the law, that is, all liars and non-cooperators have to treated the same way.

Suppose that the Gestapo, due to the the equilibria above, end up doing the random search. In this case, we can maximize the cost to the Gestapo AND maximize the number of jews being saved by ensuring nobody cooperates. If some people were truthful, then they would not be punished, and the Gestapo would have a lower cost. Of course, in a strictly utilitarian sense, this equilibrium only works if the cost to the Gestapo punishing everybody is higher than the gain in utility of some people not being punished.

Anyway, this seems very far fetched. I think in such a case, utilitarianism has to go to extraordinary lengths to prefer an equilibrium which is intuitively obvious.

### Instrumentality of truth

In our little series on truth and utilitarianism we are gradually beginning to see that utilitarians’ relationship to truth is rather complicated (see [here](https://diomavro.wordpress.com/2021/04/29/utilitarianism-is-blue-pilled/), [here](https://diomavro.wordpress.com/2021/04/13/utilitarians-truth-and-the-gestapo/), and [here](https://diomavro.wordpress.com/2021/04/08/utilitarians-and-lying/)). Today the example is more general, it applies not only to utilitarianism but any consequentialist theory which views truth as instrumental.

*Let’s say that one of your clients has died. You have two wills he left behind. One will he wrote when he was 50, he said he is utilitarian and wants to give all his assets to the most effective charity. The other will, he wrote when he was 60, he changed his mind and wants to give all his assets to his middle class family. The family is not aware he wrote two wills, they only know he has at least one.*

**What would an ethical lawyer do?**

It seems like a utilitarian lawyer would, in fact, NOT respect the wishes of the dead and simply show the family the first will. This is the case for almost all ethical systems which can be seen as optimization problems. Indeed, as long as one doesn’t have some kind of direct value for truth telling OR property rights seen as moral in themselves (or at least, outside the moral system), it seems there is no reason to give the true will. This follows from the fact that all systems which view morality as optimization problems, [make themselves morally obligatory](https://diomavro.wordpress.com/2021/03/22/utilitarianism-everything-is-obligatory/).

## Justice

### Utilitarians judge

One possible conception of law is to think of it as a long series of IF loops in programming: ‘If A and B then X, If A and NOT B, then Y… etc’. That is, the law punishes people according to circumstances. Usually, the circumstances are NOT to do with the result of the ruling. For instance, the judge should not take into account the popularity of a decision because that would make him partial. A judge does however have a significant ability to influence punishment. Let us take the case where somebody accidentally ran over another person while driving. The specific situation is this:

*George was driving on the highway at midnight, at a speed of 60 miles per hour(legal), 50 miles away from the closest city. Unfortunately, George ran over somebody that night, Sam. Sam walked in the middle of the highway and George was too slow to react. George is now sitting in court awaiting judgment for the death of Sam.*

Ordinarily, the judge would give a small penalty to George as it doesn’t seem like he could reasonably do anything to avoid it. However, a rather angry mob has emerged in front of the courthouse. They claim that George is to blame and should be executed. They claim that there is a systemic problem of Georgism, where people like George get away with killing Sams of the world. The mob has made a credible claim to bomb and kill dozens of people, should their demands not be met.

The ideal classical judge is fairly straightforward. He will look at the precedent, see how the case matches precedent or written law, and decide. Using that standard he lets people like George off the hook. The judge may sometimes give a comment or two to explain to the public why he decided to do what he did, but in general, he doesn’t care about public opinion. Of course, the judge has to think the rule is in alignment with some kind of common good, but let’s not insist on that since we may get into the weeds of legal arguments.

But now suppose that the judge is a utilitarian. The judge will reason that if he penalizes George, he will create a bad precedent that does not maximize utility. However, he also has to worry about the disutility of the threats outside the courtroom. The judge comes up with a solution. He will give George the death penalty but will cite a lot of specific reasons for it. “This decision is being made because George was driving after midnight, he was driving a truck which is more dangerous and harder to stop, George had not slept for 10 hours which made him tired, the trip wasn’t necessary, he could have stayed at his friend’s house who offered him to stay…”. Basically, he will give a list of reasons which will make events like the one that just happened, occur once in 100 years.

He reasons the disutility of this bad precedent is very low since it will only punish people once every 100 years whilst, it will possibly save 100 people now. ‘Anyway, by that time, probably some other judge will overrule my precedent so I don’t have to worry about long-term consequences.’ He then proceeds to execute George. Note: I am assuming that the Judge thinks that if a similar situation arises in the future, protesters will not be outside, if they WERE to be outside again, then the judge is even more motivated to set precedent in this way!

So it seems to me like utilitarian judges would be a disaster, this isn’t about retribution or restoration, this isn’t about levers of control or the ability of somebody to interact with their environment, it is simply taking into account the pleasures and pains of society. The general idea is that you can always make NARROW precedent.

# The entailments of utilitarian’s

### Eugenics

Most people who know me, know that I am quite staunchly anti-utilitarian. I have a few reasons for this and I plan to give them all at some point but today let me simply give one. It seems to me that the most consistent utilitarians have no choice but to be Eugenicists. I salute Singer for being true to [his](https://www.salon.com/2001/06/25/singer_2/)position. I only wish more people would advertise this implication of utilitarianism. He of course tries to stick to disabilities when he discusses this, but in his book, ‘Should the baby live’ in the first 20 pages he makes clear that even moderately disabled children should be aborted, even if there is a family ready to adopt them. Let us give a general argument:

1. All finite resources can be allocated to X or Y
2. We ought to choose between X or Y by the measure M
3. ==>X has a higher M, therefore we ought to allocate all finite resources to X.

Now we need only substitute: X = genes that have the best pleasure-pain tradeoff, Y = genes that don’t have the best pleasure-pain tradeoff, M = of higher pleasure-pain tradeoff. It seems pretty straightfoward if you ask me. But let us articulate a few things anyway just for completion. I am aware this is strictly not a valid proposition so I will include one in the bottom for the nitpicky.

So the argument does not say that all resources are finite. Indeed some could be in abundance, it simply says that IF there exist resources that are finite, those resources should be used to support the best by the measure M. This works pretty straightforwardly with utilitarianism but it could work with other things, says the person who wishes to maximize output. In general, anybody who views ethics as an optimization problem is automatically a eugenicist.

It should also be clear that the disagreement between utilitarians and a sort of Darwinism is only marginal. They both structure their argument in the same way, they might simply selecting different subsets from the population. This parallel is sometimes not obvious because utilitarianism is presented statically, but a coherent utilitarian understand that it is the long-term total utility that must be maximized.

Notice also that this is about the tradeoff. If some genes can have a great pleasure for little resources, then a utilitarian might favor numerous smaller organisms over larger ones. I suspect this is not the case since ever since Mill they might distinguish between higher pleasures and lower pleasures. So a higher consciousness can achieve greater efficiency, but if there are resources which the higher consciousness does not use, then maybe he will advocate for the smaller organisms.

In Joh Grays’s book, ‘7 types of atheism’, he puts light on how odd a phenomenon it is that two people as different as Michel Onfray and Sam Harris can converge on a single ideology. I would say that almost all non-religious scientists are of this persuasion. There very few ways to justify temporary pain and destruction to achieve a greater goal other than utilitarianism.

I don’t want to pull a Foucault here but the glove fits nicely: scientific types are merely trying to maximize their own power. First, a scientist is someone who spends their life studying IS and not the OUGHT. If scientists want to maximize their influence, they must inflate the importance of IS statements and deflate the importance of OUGHT statements. Otherwise, the scientist king would not be a good ruler. So the simplest way to deflate the OUGHT is to claim there is but ONE ought and to claim that it is so simple that they know it(utilitarianism fits perfectly).

So if one claims that only the material exists, and there is a certain material composition that is more desirable than another. They are unhindered in their attempts to reshape the world. There is no room in this mindset for admitting numerous desirable states and making morality an open-ended process, it instantly deflates the authority of the scientist, the scientists cannot make recommendations if there are numerous competing ought values.

So if somebody asks you, why you are not a utilitarian, I think it is straightfoward (but rather tongue in cheek) to answer with: ‘Because I am not a eugenicist’.

edit, for the nitpickers:

1. We ought to choose how to allocate finite resources by the higher M
2. X has a higher M
3. ==>All finite resources should be allocated to X.

### Value hierarchy

If someone has a lack of strict hierarchy in values they may exhibit seemingly inconsistent behavior. Take for example, the case where somebody has to choose between three objects, {a,b,c}. It so happens that the best one in value 1 is “a”, on the other hand the best one in value 2 is “b”.

So to break the tie, the decision maker chooses a third value, value 3 and applies it to the winners of the first round, {a,b}. It turns out that “a” is the best in value 3 so the decision maker chooses “a”.

Now suppose that the choice was between {a,c}. Once again, “a” comes out on top for value 1 but this time, “c” comes out on top for value 2. So once again the decision maker breaks the tie by applying value 3, this time “c” comes out on top.

Conclusion: If “b” is available, the choice is “a”, if “b” is not available the choice is “c”.

### Speciesm is bad

The argument is as follows: If

1) ALL human beings have a set of characteristics, X and

2) At least one animal has one characteristic that is also in X

èTherefore X cannot be used to demarcate between animals and humans.

X here can be used to represent characteristic. For example, we may say that X is the “ability to feel pain”. In that case we can say that Dolphins also feel pain, therefore pain cannot be used as a criteria to discriminate between animals and humans. This can also work with two characteristics, say the set could be {ability to feel pain, swim}.

It can also work in the reverse manner. Say that X is the ability to reason, one could try and claim that chimps may be able to reason but they cannot reason as well as humans. But then the proponent of the argument can argue,  “the best reasoning chimp can reason better than the worst reasoning human” and then claim that using X implies we should treat some chimps better than some humans.

Of course many people will bite the bullet on this accusation. They will say,” well… if that chimp has it and that human does not, then that chimp should not be treated differently than humans”. An alternative method may be that they will select X in such a way as to exclude X chimps, so if the best chimp can reason at 5/10 and the best human at a 4/10, they will set the standard at 6/10. This leads to a somewhat absurd conclusion that which humans get to be treated as humans depends on the performance of animals.

Let us give another example: Suppose that some woman is trying to decide who to choose as her husband. She could pick factors like kindness, clever, etc. I think those are less controversial, but suppose she chose “kinder than average”. In other words, the standard does not depend on the intrinsic characteristics of the person but on his ranking in the population.

This may not appear to be common but quite a few standards would implicitly have ranking criteria. Of course one could argue that the criteria “has a yacht” isn’t inherently a ranking criterion, maybe there is just a strong preference for Yachts, but it seems to me that most people understand that Yacht is the new “car”, and standards evolving in the way they do, imply a ranking.

Most people have an intuition that having a standard which is set FOR the purpose of discriminating is deeply unjust, especially to the humans who happen not to meet it. Some people want to claim that the reason we have this intuition is because we are against hierarchy of any form. Others want to argue that we seek equality and a hierarchical standard causes people on the lower end to be harmed.

So if we can’t set the standard for that purpose what is the standard for? One plausible answer is that we have already decided how to discriminate without the standard, and the standard is merely an attempt at an articulation of why we have discriminated.

**The Disney factor**

Going beyond the accusations that the argument makes, the argument itself is flimsy for the reason that it begs the question that it is a specific characteristic which demarcates between species. Just because dolphins have characteristic A and apes have characteristic B, humans can still be the only beings humans who have characteristic A and B. Of course if that isn’t enough, it probably is quite easy to come up with a characteristic that only humans have, for example contemplating God or contemplating morality or even just the ability to imagine or tell a story. In other words we can appeal to “personhood”, it just so happens that the only beings who are persons are humans.

We can see the intuitive appeal of personhood; we can call it the Disney factor. Many westerners grew up watching animations of different species that are designed to create empathy from the viewer. It is natural that fans of such animation would intuitively create a correspondence between those animals and animals in the real world. Where they are right is that IF the animals in Disney movie were real, it would not be acceptable to kill them. This seems to me to be a vision shared by meat eaters and non-meat eaters alike. That is, if it were true that an ape of the kind found in Tarzan, that can exhibit moral agency then we would obviously extend our moral code to them. Similarly, if a Superman (an alien from another planet) came on the planet, and he exhibited all the same contemplative capacities as humans have, then we would also extend our morality to him.

So it seems clear that the argument, fails, or at least if it did not fail, the argument would apply to specific non-human species (such as the apes in Tarzan or Superman) and expand our moral circle.

### Duties

As mentioned [last time](https://diomavro.wordpress.com/2021/03/22/utilitarianism-everything-is-obligatory/), if we have a given situation, utilitarianism entails that there will be one right answer about the right thing to do, in other words, everything will be morally obligatory, which leaves no room for character, today I want to highlight an entailment of this problem.

The utilitarian often asks you to take decisions as if you are in a spaceship disconnected to anything. If for instance we are asked whether one should save one baby or ten babies, the utilitarian would say, save the ten. But there is no obvious way in which they revise their views should relationships emerge, for instance, what if you are the mother of the one baby?

Classical tradition would say that a mother has a duty to her child, and though this duty may not lead to a moral obligation to save her child, it at the very least makes saving her child, morally permissible. But a utilitarian does not care, the prescription of a utilitarian is about total utility, and hence no relationships can be used as a criterion.

A utilitarian may attempt to revive duties by using something called ‘rule utilitarianism’, which means follow the rules that maximize utility, but really this now shifts the morality to plauisibly rule utilitarianism out. The rule ‘we ought to accept the rule that maximizes utility’ may actually imply that we ought NOT to be rule utilitarians. But this actually gives the whole game away, when we are actually arguing about ethics, we are arguing about what rule set or source to use when making moral decisions. Saying that we should adopt the rules that give the best consequences is only SLIGHTLY refining the set of rules. Instead of being ‘let us pick the best ethical system’, it is now ‘let us pick the ethical system with the best consequences’ but this may get you to a more or less ANY ethical system. A deontologist may be a rule utilitarian!

### Utilitarians and the urbans

A rather famous example of utilitarian reasoning is: A missile launched from Russia is about to hit a city of a million people. The missile can be redirected to hit an Amish enclave where it will wipe out ten thousand Amish.

The utilitarian says simply that one ought to maximize the number of lives saved here. They can even make a subcase that if the Amish are not wiped out, their land can increase total utility by being given to more productive hands. They can also try to justify it by saying that the Amish will in fact not be wiped out and hence can replace those who died faster due to their higher fertility. Either way, the utilitarian conclusion is fairly clear.

The utilitarian may try to argue against this by saying that city folk harm total utility because they contribute more to climate change or some other global negative effect. Is it really a surprise that the ideology that was made in the urban culture, somehow always defends the urban culture, even to the expense of non-urbans?

An egalitarian might take a different position. An egalitarian would say that each person must have equal probability to survive the missile. As such there should be a coin filp to decide if the missile will hit the Amish or the city people.

In my mind, the concept that is of importance is a sort of skin in the game. Those who benefited the most from the missile generating process should have the higher probability of being anhilited. The more independent a community is, the less it benefits from the kind of process that would result in a missile being launched.

To clarify, since the Amish are unlikely to have behaviors that cause the missile launch, it seems to me that they have a right to be less affected by it. People who deal in rhetoric against Russia or people who say contribute to cultural hegemony(which would make Russia want to respond), are more important in causing the war to take place.

To take another example, suppose Brazil somehow had a massive food shortage and they could only fix it by destroying the tribes that are within the rainforest(some of which have been around for millenia). To me, it seems like this would be unjustified, even if a million people should die, the few thousands have no obligation to destroy themselves for the million.

I would say this applies to democracy too. What if instead there was a vote, and the vote had ONE area where 90% voted to go to war with Russia and another province had 90% vote against going to war with Russia. It turned out that the 90% province had 1 million people, whilst the 90% anti war province had ten thousand people, so those who voted for the war, won. Though democracies often rely on making sure people are NOT blamed for this vote, this seems to me to have moral weight, aggregating simply isn’t morally justified.

### Blue pilled

Since utilitarianism is purely concerned with mental states, it is indifferent to the means by which those mental states are attained. As previously [mentioned](https://diomavro.wordpress.com/2021/04/13/utilitarians-truth-and-the-gestapo/), this is the inherent relationship of utilitarianism to the [truth](https://diomavro.wordpress.com/2021/04/08/utilitarians-and-lying/). The implication of this is that there is no harm to tricking people to believe things that make them happier. Or indeed, even to destroy their senses so that they can experience more pleasure.

The most famous example of this is Robert Nozick’s experience machine thought experiment in ‘Anarchy, State, and Utopia’. The idea is quite simple if it is true that one wants to maximize happiness, then the optimal solution from a utilitarian point of view is simply to enter all humans into the best video game possible for all time.

This video game would simply be perfect. If achieving your dreams would make you happiest, then the video game would have you experience your dreams as if they were real. If on the other hand, dreams are aspirations that would not make you happy if realized, then the experience machine would simply float you around those dreams. Needless to say, the machine would simply allow you to attain the highest form of pleasure possible.

Once one realizes the utilitarian doesn’t care about manipulated mental states, odd rabbit holes open up. For instance, suppose somebody is about to die, but at the last moment, we able to take their brain out and put in a vat. We are able to stimulate the brain with very cheap doses of morphine for a very long time. Is it good from the utilitarian point of view to create a farm filled with brains in vats? If it is good then it is also [morally](https://diomavro.wordpress.com/2021/04/23/utilitarianism-and-character/)[obligatory](https://diomavro.wordpress.com/2021/03/22/utilitarianism-everything-is-obligatory/)to create these large farms with brains in vats! It is also not clear to me how utilitarians value digital beings. For instance, if somebody is uploaded to a computer since there is no brain experiencing pain or pleasure, does that mean they are not to be counted? I presume the utilitarians, in fact, do not count digital pleasure, but it isn’t clear to me.

In other words, all it takes to NOT be a utilitarian is to simply say you can rather live and act in the real world, even if it means lower total pleasure. There is almost no way around this objection as far as I am aware, the utilitarians will simply claim that it IS true that we should prefer the machine to the real world. Needless to say, the utilitarians would choose the blue pill.

### Defining Human

I have been sitting on [this](https://www.degruyter.com/document/doi/10.1515/jso-2014-0037/html)paper for a year. It seems to combine a few of my interests. Ontology seems to have a developed language for talking about emergence. Some things, are not reducible to their components. If you can explain X by going to a different scale then X is not an ontological thing. From the paper:

*In order to be in ontology at time t, a concrete entity, property, or primary kind must be (ontologically)****[2]****irreducible and ineliminable without loss of completeness (at time t). An item is irreducible and ineliminable (at time t) if and only if it is not reducible and not eliminable (at time t). An entity (or property) is reducible if and only if it is entailed by local microphysical properties. More colloquially, an entity (or property) is reducible if and only if it is “really something else” (Fodor 1987, p. 97). An entity (or property) is eliminable if and only if a complete ontology does not entail that it exists (or is instantiated). Primary kinds belong in ontology, because kinds that are reducible or eliminable are not primary kinds.****[3]***

*Social ontology, as I construe it, is part of ontology simpliciter. As part of ontology, social ontology should include all social entities, social kinds and social properties that are irreducible and ineliminable. A social property is one for which social or linguistic communities are necessary for its instantiation. A community is one whose members bear significant intentional relations to one another. I say ‘significant’ to rule out an aggregate of people waiting to cross the street as a community. They may all happen to have curly hair, but they are not a community. A community is a group of persons with a measure of cohesion, with common intentional properties or relations such as shared interests or values or language. A social property belongs in social ontology if and only if it is irreducible and ineliminable. An entity is social entity if and only if it has a social property, and a kind is a social kind if and only if it has social entities as members.*

One can have a first person perspective 1) in the nonconceptual sense(rudimentary) or 2) they can have it on the conceptual sense(robust). For humans, they always have the capacity to develop the second sense, even if they never end up developing it. Animals can only arrive at the rudimentary sense, humans generally pass by the rudimentary sense.

*A robust first-person perspective – an ability, not just to recognize oneself as distinct from everything else, but to*conceive*of oneself as oneself in the first person – makes possible almost all our characteristic human activities. To name a few: One can deliberate about what to do and can attempt to rank preferences and goals, and try to resolve conflicts among them (and thus is a rational agent); one can reflect on her motives; one can have a life of moral significance; one can have an inner life; one can conceive of herself as having a past, some of which is accessible to memory, and as having a future, some of which is accessible to intention. Robust first-person perspectives enable us to realize that we are agents, to take responsibility for things that we do, to recognize that we are subjects of experience, to care about the future, to change our habits in light of rational assessment of our goals. These abilities – made possible by our robust first-person perspectives – are unique (as far as we know) in the universe (Baker 2000, p. 147–164; Baker 2011).*

*We have ample linguistic evidence of first-person perspectives in the use of first-person pronouns embedded in first-person sentences whose main verbs are linguistic or psychological verbs – ‘I am glad that I have such good friends’, ‘I wish that I were not in pain,’ ‘I told you that I was in pain’, and so on. If I think or say that I am glad that I have such good friends, I am conceiving of myself from the first-person, without needing a name, description or other third-person referential device to refer to myself. I’ll call such complex first-person thoughts ‘I\*-thoughts’.*

*There are a couple of features of I\*-thoughts worth noticing: First, they are not limited to “Cartesian” thoughts about what one is thinking; they include mundane thoughts like “I wish that I\* were in the movies”. Second, I\*-thoughts need no recourse to any peculiar object like a self, or a soul, or an ego. My I\*-thoughts refer to me, a person – the same entity that you refer to by saying, “Lynne Baker”. There is no special object, distinct from a whole embodied person, that is a self. All I\*-thoughts are manifestations of robust first-person perspectives. Our ability to think such thoughts depends crucially on language.*

These can be used as an attibute to play the name the trait game for moral value. Once you get your head around it, it makes sense but it isn’t neccesarily easy to express in a conversation. Nevertheless even if we do use this as a clean categorization, it will fail to capture your moral system because of the [‘superior alien’](https://diomavro.wordpress.com/2021/05/04/utilitarians-and-humanity/) objection. Which is why, at some point, you have to give up on utilitarianism.

### Humanity is second

Suppose an alien species has landed on our planet. Let’s say that they are just like humans but they lack a certain pain receptor. Instead they have a robotic reflex which has programmed them to react quickly whenever they are damaging themselves. So in expectation they are just like humans but feel less pain. These aliens find that humans are simply wasted space on this planet so they will simply kill all humans and place themselves on earth.

For some reason, a utilitarian human is going to decide who wins the war. Without hesitation, it seems like a utilitarian would destroy humanity, themselves included. This is merely a consequence of valuing pleasure states over anything intrinsic to humanity. If you take an impartial point of view, you are in fact putting humans second.

It seems like utilitarians think the function of morality is to make a better universe. In contrast, most non-utilitarians think the function of morality is make humans live in harmony with each other, their environment or the will of God. It is good to try and generalize morality so that it can take into account more situations but there is no point in generalizing if the result is a loss of function. This is also what Nietzsche tried to point out.

It seems that in order to keep morality functional for humans we must keep some of the constraints that prevents it from being too universal.

note to self: Perhaps it’s worth doing a patriotic example too.

### Particularism and Singer

The idea that we can reduce ethics to a single principle is absurd in many ways but I want to give a brief few comments on Singer. Singer has spent a good chunk of his career defending preference utilitarianism and then switched to classical (or hedonic) utilitarianism.

What is quite stupefying to me is what his reasoning process for changing must have been. Presumably, people kept giving him examples of preference utilitarianism that made him uncomfortable, or perhaps he was trying to find strategies to defend his beliefs and found that classical utilitarianism actually worked better.

But clearly for this whole process to play out it must be that he held at least some of his views constant! That is, he held the view that abortion is justified and animal suffering is not and was looking for the principles to justify his view. We need only note that in his shift from one kind of utilitarianism to another, he did not actually change position on any of his views!

This invariance in policy prescriptions implies that the principle is merely secondary. While the attempt to reduce morality to a single principle might be fun, it is, practically speaking, not important. If you find that your principle goes against your intuition, then it is not your intuition which must be dropped, but the principle.

This reminds me James C Scott’s Seeing like a state. There is an image where before property was drawn in strange unpredictable patterns, then, similar to what has been done in Africa, property was arbitrarily re-drawn in straight lines, presumably because it was easier to bureaucratically process. This is analogous to modernity and past customs. Past customs draw the line between what is permissible and non-permissible in strange complex ways, and the modernist comes to try and draw straight lines and categories.