

5. Varieties of consequentialism

Martín Abreu Zavaleta

July 13-14, 2015

We've been examining Aristotle's view about the nature of the good. According to him, the highest good that a person can aspire to is a life of action in accordance with virtue. He thinks that this is just what it is to be a good person.

However, we also noticed that this way of thinking about what it is to be a good person may not match some of the ordinary usage. Oftentimes, when people talk about being a good person, they talk about someone who does good deeds. We are about to consider views that, in that sense, seem to be more in line with ordinary usage. Those views are not so much interested in discovering the highest good for a person. Instead, they are interested in finding out what are the *morally good* actions. In this contemporary sense, to be a good person is to be someone who does morally good actions.

The two senses are close to each other, but there are also significant differences between them. One similarity between them is that they both think that the good person does good deeds. For instance, Aristotle thought that a virtuous person would act in accordance with virtue, i.e. do virtuous things.

Though Aristotle did think that the virtuous person did virtuous deeds, he also thought that such deeds were not virtuous because of the consequences they had; rather, he thought that they were virtuous partly because they were performed by a virtuous person. Two actions may be alike in all visible respects, yet only one of them may be virtuous. This will depend on which of the two is made knowing that the action is in accordance with virtue, wanting the action to be in accordance with virtue, and doing the action because of some stable disposition to do actions in accordance with virtue.

The views of the good that we'll consider now care mostly about the "visible" properties of actions. In particular, they care mostly about their consequences. For that reason, these views are called *consequentialist*. Note that the focus of our inquiry has shifted from what is good for a person to what constitutes a good action. Virtue and the function of a person don't play any role in these new views.

Moreover, Aristotle did his best to resist the idea that there was a rule, or a set of necessary and sufficient conditions, that determined what to do in each case. This is why he says:

[T]he account of particular cases is more inexact. For these fall under no craft or profession; the agents themselves must consider in each case what the opportune action is, as doctors and navigators do. The account we offer, then, in our present inquiry is of this inexact sort; still, we must try to offer help. (1104a5-13)

Consequentialists, on the other hand, aspire to give necessary and sufficient conditions for the action to be good. In principle, this allows them to determine a procedure for deciding what to do in

each particular situation. However, as we shall see, such decision procedures are very difficult to implement in actual practice.¹

1 Basic commitments of Consequentialism

According to consequentialists, the normative properties of an action—whether it’s right or wrong, good or bad—depend only on their consequences. In a past session we considered the case of someone who did actions with nice consequences, but who didn’t want to perform those actions, and in fact hated those kind of actions. According to consequentialism, all those features about the agent’s mental state are irrelevant to our assessment whether the action was right or wrong. All that matters is that they have good consequences.

Different strands of consequentialism may think that different consequences are desirable. For instance, classical *utilitarianism* claims that an act is good if it maximizes happiness. Utilitarianism understand happiness in a very minimal sense, as having *pleasurable sensations*. However, not all consequentialists think that happiness is the thing to be maximized. Others think that *welfare*, *desire satisfaction*, among others, are the things that we should try to maximize with our actions.

Consequentialist theories can vary in many respects. Here we’ll discuss only four such dimensions: which thing should we maximize? what kind of consequences do we care about? should we care about the consequences of particular actions, or of actions that follow a certain rule? consequences for whom?

2 What kind of thing should we maximize?

As I already mentioned, classical versions of consequentialism were, for the most part, versions of *act utilitarianism*. On such views, a good action is one that maximizes utility. We can ask different questions about this definition, but in this section we’ll focus on how exactly to understand the notion of utility.

Roughly, utility is usually understood as *pleasure* of some kind or other. Unfortunately, this still isn’t precise enough. As we saw in our discussion of Aristotle, there may be different kinds of pleasures. If this is so, we can still wonder which kinds of pleasures consequentialists care about.

One straightforward way of defining pleasure is simply as a *pleasurable sensation*. This is the kind of pleasure that you get, for instance, when your arm feels itchy and you scratch it. The sensation of relief that you get from scratching in that kind of case is a pleasurable sensation. Or think about the pleasure you get when you eat delicious food. Those sensations also count as pleasurable sensations.

Unfortunately, this definition doesn’t seem to leave a lot of room for other kinds of pleasure. On this very restricting definition, the pleasure you get from reading a good book, watching a good ballet performance or listening to beautifully composed and executed music doesn’t count as a pleasurable sensation. Though we enjoy those things partly because of the sensations they produce, it seems unlikely that the pleasure they produce in us can be reduced simply to a sort of brute sensation like taste, sight, or touch. We usually think that the pleasures coming from our interactions with the art are of a more intellectual sort.

¹Much of the content of the present notes is taken from the Stanford Encyclopedia of Philosophy on Consequentialism and lectures from Samuel Scheffler’s Spring 2015 course on Political Philosophy.

Once we plug in the simple definition of pleasure into our definition of act utilitarianism, we get a view that predicts that you should prefer scratching your arm when it feels itchy instead of watching a play or going to a concert. This doesn't seem right. So it seems that any plausible version of utilitarianism will have to use a more comprehensive definition of pleasure, one that reflects our intuition that watching a play is more valuable than scratching our arm when it feels itchy.

This more comprehensive kind of pleasure will include things like *being pleased that our friends love us*, or *being pleased to have found an affordable place to live*. Notice, however, that this may let too many pleasures in. For instance, *being pleased that a hundred people are being tortured* or *being pleased that people die in the Middle East*. Many of us wouldn't think that acting in ways that would maximize the latter kind of pleasure would be right. In fact, we would think that it's wrong to act in ways that maximize those sorts of pleasures.

Perhaps these sorts of problems could be solved by working even more on the right definition of pleasure [**Question:** What do you think the act utilitarian could say in response to these objections?]. However, it might be better to consider a different line of objection.

Recall Nozick's thought experiment involving the experience machine. In it, we are invited to consider the possibility of plugging ourselves into a device that would produce in us whatever experiences we want to have, presumably with the option of getting only pleasurable experiences in some sense amenable to the utilitarian's purposes. Nozick hopes that most of us would prefer not to be plugged into the machine. Furthermore, he thinks this shows that we value things over and above pleasurable experiences. The point is that, in addition to pleasure, we value our contact with the real world and whatever that implies. Thus, even if we somehow managed to maximize utility in the way the utilitarian advises, we would be missing out other forms of value. Thus, acts that maximize pleasure are not only the right acts, because there may be other valuable things that such acts miss and would be better to do.

Other varieties of consequentialism claim that we should maximize the satisfaction of people's desires or preferences, welfare, among other things. Notice that these versions of consequentialism may not be as easy a prey to Nozick's observations—which is not to say that Nozick's observations are effective against classical utilitarianism in the first place. Welfare and satisfaction of desires require something more than just having certain experiences or being in certain mental states. For someone's wish to pass the class to be satisfied, it's not enough that she has an experience as if she had passed the class. For her desire to be satisfied, she has to *actually* have passed. So it's not so easy for Nozick's original thought experiment in that respect. **Question:** How could Nozick refine his point against welfarist or preference-maximizing versions of consequentialism?

Another way in which consequentialists may disagree among themselves is with respect to the right way to understand the notion of *maximization* and whose good we should maximize. We'll consider this issue in the next section.

3 How to understand maximization?

As we have stated repeatedly, consequentialist views equate good actions with actions that maximize some thing. But how exactly should we understand maximization? For the sake of simplicity, let's suppose that the thing to be maximized is utility, understood as some sort of pleasure. Given this choice, we can ask the further question: whose utility should an action maximize? Should a good action maximize the utility of the person who performs it (i.e. the agent)? Should it maximize the utility of the people in the close vicinity of the agent, or of the people that the agent cares about?

Perhaps it should maximize the utility of people without restriction, or even the utility in the world as a whole (including animals, if they too can have utility). Even once we fix on an answer to these questions, we still have options. For instance, suppose that we have decided that the right thing to do is to maximize the utility in the world as a whole. Still, we can ask if we should maximize such utility *in the present*, in the *near future*, or even just the utility in the world across all times.

As you can see, deciding whose utility one's actions ought to maximize is not easy. If we start with the more particular things (e.g. an agent should always act so as to maximize her own utility), we soon run into troubles: imagine someone who really likes torturing people. Surely torturing people would be wrong regardless of how much joy it brings to the torturer, yet a version of consequentialism on which the right thing for an agent to do is to maximize her own utility will claim that the right thing to do for the torturer is to torture people. This consequence is unacceptable.

Perhaps the right thing to do is to maximize the utility of the people that we care about, or of the people in our close vicinity. But what is so special about those people? Consider a person you really care about, call her the close person. Perhaps this close person is one of your parents, your girlfriend or boyfriend, or a close friend. You have \$100 that you can spend in making a gift that will make the close person happy, or you can spend them in buying four nets that will protect children in Africa from infection-transmitting mosquitoes, potentially saving their lives. Surely there is more utility in a whole lifetime than in whatever joy the close person would get from the gift that you make to her. It is clear (let's suppose) that the four recipients of mosquito-nets would gain much more utility from the \$100 you would spend than the close person would get from your gift. If the version of consequentialism that we are considering is right, however, what you ought to do is buy the gift for the close person. But this raises an important question: what is so special about your close friend that the utility she would get from your gift is more important than all the utility the other four people would get from the mosquito-nets?

In most cases, the answer will be that there is nothing special about the close person that would justify such preference. Because of this kind of considerations, it seems we are forced to endorse versions of consequentialism with broader and broader scopes. However, this too leads to counter-intuitive consequences. We can use the reasoning above to reach the conclusion that there is nothing special about the utility of present people, or even about the utility of people. If, for instance, it turns out that the world as a whole would have much more utility if there were no humans (e.g. because animals would live long and happy lives or something like that), this last version of consequentialism would claim that we ought to painlessly kill all humans, or perhaps just prevent humans from further reproduction until, eventually, we all die. But again, this seems to be unacceptable.

Question: Do you think that there is a "right" scope of people or beings the maximization of whose utility should be targeted by our actions? Which beings (people) would those be, and why do you think their utility should be targeted above the utility of others? Perhaps that group is different for each agent, or perhaps it isn't.

Notice that the problem is aggravated by consideration of future people. Perhaps if we sacrifice the utility of present people, we will maximize the utility of future people in a way that would bring about an overall state with more utility than if we just tried to maximize the utility of present people.

Further precisifications of maximization

Suppose we have decided on the group of people relevant to our calculations of utility. There is still some work to do with respect to refining the notion of maximization. For there are at least two ways

of understanding maximization. On one version, what we care about is maximizing the *total* net amount of utility among the population that we care about. On another, we care about the *average* utility per member of the population.

In many cases, it may turn out that what maximizes total utility also maximizes average utility, but things don't have to be that way. Consider two populations A and B. Population A has a million members. Because of its size, each of its members has just enough resources to live a life producing net positive utility. Let's suppose we can measure utility using real numbers, and let's call the units of utility *utils*.² Let's say that each member of A has throughout her life a total of 1 util. So population A produces a total of 1 million utils.

Population B is comparatively smaller, with 100,000 members. However, the small size of the populations allows them to get more utility from the resources they have, so that each person produces 4 utils throughout her life. Thus, population B produces less than a half million utils, much less than the total amount of utility produced by population A.

According to *total utilitarianism*, if we had the opportunity to choose which of these two populations we can create, we should choose to create population A; according to *average utilitarianism*, we should choose to create population B. In fact, total utilitarianism seems to have the consequence that if we have the chance to create a population (e.g. by means of certain government policies) in which everyone had only a minimal amount of utility, but that was so big that the total utility was more than the total utility we actually produce, then that's what we ought to do. Some people call this *the repugnant conclusion*. It is repugnant because it tells us that we should produce many people whose lives are barely worth living instead of few people with really good lives. **Question:** which of the two things should we do, if any?

4 Which kinds of consequences?

Nothing that we've said so far tells us which kinds of consequences we should care about. One option is to take the morally right actions to be those that *actually* maximize good consequences. Another is to think that the morally right actions are those that maximize the *expected* good consequences.

To illustrate the difference, suppose that you only have imperfect information and there are two courses of action available to you. The first course of action has three possible outcomes: there's a 0.1 chance that it will produce 1000 utils, a 0.5 chance that it will produce -100 utils, and a 0.4 chance that it will produce 10 utils. The second course of action also has three possible outcomes: there's a 0.3 chance that it will produce 100 utils, a 0.5 chance that it will produce 60 utils, and a 0.1 chance that it will produce -10 utils.

The expected utility of a course of action is the sum of the utility of every possible outcome, each multiplied by the probability that it obtains. So the first course of action has an expected utility of $100 - 50 + 4$, that is 54 utils. The second course of action has an expected utility of $30 + 30 - 1$, that is 59 utils. So the expected utility of the second course of action is higher than the expected utility of the first course of action. Now suppose that, as a matter of fact, if you opt for the first course of actions, and despite its low probability, your action will produce 1000 utils. Which is the

²We don't need to care too much about what this means for now. Think of utils merely as an useful device for comparing different versions of utilitarianism. Presumably, similar units can be devised for versions of consequentialism that don't take utility to be the primary good.

right course of action?

According to the version of consequentialism that focuses on maximizing *actual* good consequences, you should pursue the first course of action. According to the version that focuses on maximizing *expected* good consequences, you should pursue the second one. **Questions:** Which of these views is better, and why? What are the advantages and disadvantages of each view? Can you think of other objections to these versions of consequentialism? What would happen if there was a person that obtained a lot of utility of the misery of others?

Some people think that the right answer is close to this: the version that focuses on *actual* outcomes is better as an account of the nature of good actions, but the version that focuses on *expected* outcomes is better as a decision procedure. On this view, for an act to be morally good, or for it to be the morally right thing to do, is for that act to maximize actual utility. However, we rarely know which particular outcomes our actions will have. In the absence of perfect information, the most rational thing to do is to choose the course of action that maximizes *expected* utility. On this view, we won't always do the morally right thing, but at least we will be blameless because we chose the best course of action from a rational perspective.

5 Acts vs. Rules

There is yet another way in which varieties of consequentialism, and utilitarianism in particular, can differ. So far, we have taken consequentialism to assume that the things that are morally right or morally good *in the most fundamental sense* are actions. However, some versions of consequentialism disagree. We'll focus on a variety of consequentialism on which the things that are morally right in the most fundamental sense are not actions, but *action-guiding rules*. On this last view, actions can be right or wrong, but only because they are dictated by a rule that is itself good or bad.

For instance, rule-utilitarians think that the right thing to do is to act in accordance with a rule whose application maximizes utility. For instance, it may be that in some particular case giving food to a hungry person doesn't maximize her utility: perhaps the food will make her sick because she had an allergy that no one knew about. But in general, giving food to hungry people makes them happy. So the right thing to do is to give food to hungry people. Notice that the latter is the right thing to do, but not because of its intrinsic properties. It is the right thing to do because *in general*, giving food to hungry people maximizes utility.

Unlike act utilitarianism, rule utilitarianism seems to be in line with our intuitions in some important cases. One of the most famous ones is known as *Transplant*.³ In this scenario, we are asked to imagine that there are five patients in need of a transplant, but each of them needs a different organ. There is also another patient who is in the hospital merely for routine tests, and a doctor administering the tests. The doctor could maximize the overall utility in the world by killing the healthy patient and giving his organs to the five patients in need of a transplant—suppose that each of the people receiving an organ would emerge healthy from the operation, the doctor wouldn't get caught, etc.

According to act utilitarianism, the doctor ought to kill the healthy person to give his organs to the patients in need of transplants, since this is the *action* that would maximize utility. However, most of us think that this is unacceptable: it's simply wrong to kill the healthy person in this kind of situation.

³As far as I know, the case is due to Philippa Foot, but I'm not certain of this.

Rule utilitarianism can say something more in line with our intuitions about this case: in general, killing a person to give his organs to other people doesn't maximize utility. All sorts of things could go wrong, from the surgery itself to the probability that the perpetrator gets caught. So killing the healthy patient in that particular situation is not the right thing to do. In fact, it is wrong because *in general*, operating in that way would prevent a lot of utility from being produced, and would instead create a lot of *disutility*. Think about what would happen if doctors routinely killed patients: people would stop going to hospitals, and perhaps die younger or suffer from illness, etc.

6 Objections to consequentialism

One objection to consequentialism is that it imposes an idea of moral good and moral rightness that is too demanding. Samuel Scheffler, for instance, offers the following kind of consideration. Suppose that I have a serviceable but old and dirty pair of shoes. I want to buy some new shoes that cost \$100. If I spent that money buying mosquito nets for children in Africa, I would maximize utility. So if it is wrong to do anything other than what maximizes utility, it would be wrong to buy the new pair of shoes. But buying the pair of shoes doesn't seem wrong at all. Surely, it would be morally better to donate the money to charity, but doing so seems *supererogatory*, rather than demanded by morality.

This line of reasoning doesn't only apply to our money-spending habits. Think about all the good that you could be doing if instead of watching TV on a Tuesday night you were actively trying to helping others in some way. However, it doesn't seem to be wrong to stay home and watch TV.

The point of these remarks is that maximizing utility with each of our actions would require us to make significant changes in many areas of our lives. A consequentialist morality simply seems too demanding in ways that we wouldn't expect it to be. In a couple of sessions, we'll read an article by Susan Wolf in which this issue is examined in more depth.

Another kind of problem was originally raised by Philippa Foot, and presented in its canonical formulation by Judith Jarvis Thomson. Here is Thomson's canonical formulation:

Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit of a valley at that point, and the sides are steep, so you must stop the trolley if you are to avoid running the five men down. You step on the breaks, but alas they don't work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the straight track ahead. Unfortunately, ... there is one track workman on that spur of track. He can no more get off the track in time than the five can, so you will kill him if you turn the trolley onto him (Thomson 1985, 1395)

What would you do? Classical utilitarianism has a straightforward answer: veer so that you kill only one person instead of five. This may seem compelling in the case at hand, but of course there are different versions of it.

For instance, we could make an analogy between this canonical *trolley problem* and the following case, which was offered by Foot before Thomson's canonical formulation. There is a mob accusing an innocent person of having committed some crime. Suppose you are the sheriff, and you have two choices: you can protect the innocent person, but if you do that the mob will set that person's town on fire; alternatively, you can frame the innocent person and send her to jail or something

like that, in which case the mob will rest contented. What should we do? Again, consequentialism has a straightforward answer, but is that the right answer?