

Peter Singer, AC
Ira W. DeCamp Professor of Bioethics
University Center for Human Values
Princeton University

Hedonistic Utilitarianism: A Defence

Delivered at the National and Kapodistrian University of Athens on the occasion of the conferral of the degree of Doctor, *honoris causa*, May 25, 2015.

1. Introduction

Utilitarianism may well be the most influential secular ethical theory in the world today, but it is also one of the most controversial. It clashes, or is widely thought to clash, with the idea that there are absolute moral rules, or in other words, that there are some things that it is never right to do, no matter what the consequences. Dostoyevsky provides a famous example of this objection to utilitarianism in *The Brothers Karamazov* when he has Ivan asks Alyosha this question:

“Imagine that you are creating a fabric of human destiny with the object of making men happy in the end, giving them peace and rest at last, but that it was essential and inevitable to torture to death only one tiny creature -- that baby beating its breast with its fist, for instance -- and to found that edifice on its unavenged tears, would you consent to be the architect on those conditions? Tell me, and tell the truth.”
“No, I wouldn’t consent,” said Alyosha softly.

This is an entirely hypothetical example, of course, but it does point to a divide between utilitarians and those who believe that there are some absolute moral rules that must never be violated. We shall see what the utilitarian can and should say to such issues. But first I shall ask: What is utilitarianism?

Utilitarianism is a member of a broader family of theories known as consequentialism. Consequentialism is the view that the right action is the one that will bring about the best consequences. Within this framework, utilitarianism is distinguished from other forms of consequentialism by holding that “best consequences” should be understood in terms of the welfare of all those affected by the action. There are different views as to how we should understand welfare, for example in terms of the satisfaction of preferences, or the conscious experience of happiness or pleasure. I shall say more about this shortly.

2. Justifying Utilitarianism.

Can utilitarianism be justified? To ask such a question is to raise the larger question: How can any fundamental ethical principle be justified? John Stuart Mill, in his essay *Utilitarianism*, has a chapter in which he discusses “Of what sort of proof the principle of utility is susceptible.” The answers given by utilitarians range from “no proof at all” to the claim that utilitarianism rests on some self-evident truths. I will discuss six possible answers, in order of closer approximation to what I believe to be the truth.

J.J.C. Smart:

Smart was some kind of subjectivist about ethics. That is, he thought that it is not possible to offer any foundation for a normative ethical theory. In all his writings on utilitarianism, from *Outline of a System of Utilitarian Ethics* onwards, he essentially appealed to his readers' benevolence. We can paraphrase his appeal to his readers as follows: "Think about suffering. Don't you deplore it? Wouldn't you prefer a world in which there was less suffering, and more happiness? If that is your overriding goal, as it is mine, then join me in promoting utilitarianism and acting in accordance with it."

Such an appeal will, of course, do nothing to persuade those who, even though they may deplore suffering, deplore the violation of moral rules (such as "Do not torture!") even more. Smart could try to explain why having a rule against torture will generally have good consequences, because it will prevent abuse of innocent people who are suspected by police or military authorities of being in opposition to them. Nevertheless, some people may hold that this, and many other moral rules, are important for their own sake, and not only because they generally have good consequences.

Jeremy Bentham:

The founding father of English Utilitarianism claimed, at the outset of his *Introduction to the Principles of Morals and Legislation*, that terms like "right" and "wrong" only have a meaning when they are used to assess the utility of actions, and otherwise have no meaning at all. Although there is some truth to the claim that the meaning of these terms is somewhat mysterious when they are used to refer to moral principles known by divine revelation, or by our conscience, and that in contrast they can be given a clear and precise meaning when used so that "right" means "the action that, of all the actions open to me, will lead to the most utility" and "wrong" means "an action that will lead to less utility than some other action open to me," we cannot pretend that we do not understand someone who uses moral terms differently. For example, someone we may disagree with someone who says "Telling a lie is always wrong, even if by doing so one saves an innocent life," but we cannot deny that we understand such a sentence.

John Stuart Mill:

Mill, in the chapter of *Utilitarianism* on what kind of proof the principle of utility is susceptible, argued that as everyone desires his own happiness, the aggregate of what everyone desires must be the general happiness. The argument was severely criticized by G.E. Moore, in *Principia Ethica*, as committing the naturalistic fallacy, that is, the fallacy of defining "good" in terms of a natural property such as "happiness." Moore argued that if such a definition really captured the meaning of the word "good" then it would be senseless to ask "I know that this action will produce the most happiness, but is it good to do it?" Moore considered that such questions are still "open," in contrast to, for example "I know that John is a bachelor, but is he unmarried?" The latter question really is pointless, because if we know someone is a bachelor, we know that he is unmarried. The same cannot be said about the relationship between "good" and "productive of the most happiness" so this definition cannot capture the full meaning of "good". Later critics have also accused Mill of violating David Hume's injunction against deducing an "ought" from an "is".

R.M. Hare:

Hare, who was my own teacher when I was a graduate student at the University of Oxford, argued in *Freedom and Reason* that a proper analysis of the meanings of moral terms such as “ought” can show that the term carries with it the principle of universalizability. In other words, if I claim that you ought to pay your taxes, then I am committed to holding that anyone else in relevantly similar circumstances ought to pay his or her taxes. We might, of course, dispute about which circumstances are “relevantly similar” but I cannot argue that you ought to pay your taxes because you are you, whereas it is not the case that I ought to pay my taxes, because I am not you. Ultimately, in *Moral Thinking* and other late works, Hare argued that universalizability, by requiring us to put ourselves in the position of all the others affected by our actions, with the preferences which they hold, leads to a form of preference utilitarianism in which we maximize the satisfaction of preferences.

Henry Sidgwick:

Sidgwick was the last of the great nineteenth century English Utilitarians. Although less well-known than Bentham and Mill, his masterpiece, *The Methods of Ethics* is increasingly being recognized as one of the greatest works of moral philosophy ever written.

Sidgwick’s argument for utilitarianism can be seen as consisting of stages. First, he undertakes a careful critical examination of common moral intuitions in which he shows them to be lacking in precision or independent self-evidence. Thus he argues that for these principles to give adequate guidance in difficult cases, they require supplementation from the principle of utility. Second, he claims that if we imagine ourselves taking what he calls “the point of view of the universe” we can see that our own good is of no greater importance than that of any other individual. As rational beings, we ought to aim at good generally, not merely at the part of it that is my own good. This is, he maintains, a self-evident axiom of reason. In the third stage of his argument, Sidgwick argues that careful reflection on what is good enables us to see that while many things are good instrumentally, only conscious states can be good in themselves. He therefore concludes that pleasure or happiness is the only thing that is intrinsically good, and pain or suffering the only thing that is intrinsically bad.

Despite reaching this conclusion, Sidgwick was unable to conclude that hedonistic utilitarianism is the only rational ethical view, because he felt unable to deny that egoism is also rational. Both egoism and utilitarianism are forms of hedonism – the one focused on the happiness of the individual agent, the other on universal happiness – and they will sometimes conflict. This led Sidgwick to despair at ever finding a solution to what he called “the dualism of practical reason” but without such a solution he felt unable to put ethics on a rational basis.

Katarzyna de Lazari-Radek and Peter Singer:

To bring this fully up to date, I will refer to my own work, co-authored with Katarzyna de Lazari-Radek. In *The Point of View of the Universe* we draw on and generally support Sidgwick’s achievement, but we go further by offering an evolutionary debunking argument against many of our moral intuitions. We argue that we can explain these intuitions as having been formed within us by a process of evolutionary selection, which enabled us to survive and reproduce, and have surviving offspring, under the conditions under which our ancestors, both human and pre-human social mammals, lived for millions of years. Because the outcome of the evolutionary process is what is conducive to survival and reproduction under the specific circumstances in which that process took place, understanding how evolution produced our

intuitions has the effect of debunking them. Evolution conduces to survival, not to truth. So an evolutionary explanation of an intuition undermines its claim to be a moral truth or a truth of reason. This undermining applies to egoism, but it notably fails to apply to universal rational benevolence, because it is hard to see how benevolence to complete strangers, or even to animals, could assist individuals, or their offspring, to survive. Thus we argue that egoism is discredited as a rational principle, but universal benevolence remains and can provide the required foundation for utilitarianism.

3. Direct and Indirect Utilitarianism: Act Utilitarianism and Rule Utilitarianism

The standard version of utilitarianism evaluates each act in terms of its consequences, and so is known as Act Utilitarianism. Some utilitarians have suggested that we should instead evaluate rules in terms of their consequences, and acts in terms of their compliance with the best rules. This is Rule Utilitarianism. At times, John Stuart Mill appears to suggest that this is how utilitarianism should be understood, although on this, as on several other issues, he is not entirely consistent. Brad Hooker is perhaps the best-known contemporary exponent of rule utilitarianism. I defend Act Utilitarianism, but in a form that allows space for a two-level view, as defended by R.M. Hare in *Moral Thinking*. On this view, we can accept most moral rules as useful guides in everyday life, while acknowledging that in some unusual circumstances it may be better to violate them. Thus in everyday life we follow the everyday rules against lying, cheating, breaking promises, harming others, and so on, but if a would-be murderer were to knock on our door and ask us the whereabouts of his intended victim, and we knew both where the intended victim is, and that the person asking us was likely to murder her, we would lie.

4. What Should We Maximize?

As I have said, utilitarianism is the form of consequentialism that understands “best consequences” as referring to the well-being of sentient beings. But in what does “well-being” consist? Classical or hedonistic utilitarianism, as developed by Bentham, Mill and Sidgwick answers this question by understanding the highest possible “well-being” for an individual as the greatest possible surplus of pleasure over pain. In the twentieth century, under the influence of economics and the desire to make welfare something that can be observed, it became popular to understand welfare as the satisfaction of preferences. Preferences, it was thought, can be observed because our behavior when we choose reveals our preferences. For example, if I have a dollar and when shopping I choose two bananas, priced at 50 cents each, rather than one orange, priced at a dollar, this shows that I prefer two bananas to one orange. It does not show that I will get more pleasure out of eating the two bananas than I would get if I were to eat the orange. Maybe I will, but maybe I won’t – pleasure is a mental state and this has remained, at least up to now, beyond scientific observation. (The growth of neuroscience, however, means that this may be about to change.)

For many years, I was a preference utilitarian, largely because preference utilitarianism view does not impose on others a view about what their own good consists in. It allows each individual to decide. But for a variety of reasons, I have moved away from this view. In part, this is because I have been persuaded that some preferences are simply irrational, and a person’s welfare is not advanced by satisfying his or her irrational preferences. Preferences can be

inculcated in us by various means, including advertising, and yet the satisfaction of the preference may bring only momentary happiness, or no happiness at all.

Preference utilitarians have tried to meet this objection by adjusting their position so that it refers to the satisfaction of preferences that people would have, if they were well-informed and thinking calmly. But it is still possible to be well-informed and calm, and yet irrational. Derek Parfit invites us to imagine someone with what he calls Future Tuesday Indifference. This person is just like us in most ways, preferring to experience a mild discomfort, such as a slight headache, to extreme agony, like being tortured. But there is one exception: he doesn't care about agony that will happen on some future Tuesday. So, if he has to choose between a mild headache on Monday, or agony on a future Tuesday, he chooses the agony on the Tuesday. When Tuesday comes, and he is suffering agony, however, he regrets this choice, because then it is a present Tuesday, not a future Tuesday, and his attitude to agony on a present Tuesday is no different from his attitude to agony on a Monday or Wednesday, which is to say, it is no different to our attitude to agony on any future day. This seems to be clearly an irrational set of preferences. Once we acknowledge that preferences can be irrational, however, that raises the question of what preferences are rational and it is at least arguable that – as Sidgwick argued – if we were fully rational we would prefer to maximize the welfare of ourselves, and others, understood in the hedonistic sense of the greatest possible surplus of pleasure over pain. Thus a critique of the irrationality of preferences can lead us to hedonism as the best understanding of welfare, and thus to hedonistic utilitarianism.

5. Objections.

From Bentham's time onwards, utilitarianism has been thought to be vulnerable to many different objections. Against Bentham, it was argued that utilitarianism is a dangerous principle, because it undermined the traditional institutions of British society, including the power of the aristocracy. Indeed it did, because it was a reforming movement, supporting the abolition of the slave trade, universal suffrage, reform of the terrible conditions of English prisons, laws against cruelty to animals, and better treatment of the poor.

All of the leading utilitarians – Bentham, Mill, and Sidgwick – were clear that the utilitarian principle extends to nonhuman animals. Utilitarians supported the introduction, in England in the early decades of the 19th century, of the first laws to prevent cruelty to animals. Yet to some, the fact that utilitarianism implied that the welfare of animals counts equally alongside human welfare was considered an objection. William Whewell, a nineteenth century philosopher, argued: “It is ... to most persons not a tolerable doctrine, that we may sacrifice the happiness of men provided we can in that way produce an overplus of pleasure to cats, dogs, and hogs.”

To this John Stuart Mill gave a ringing response:

It is “to most persons” in the Slave States of America not a tolerable doctrine that we may sacrifice any portion of the happiness of white men for the sake of a greater amount of happiness to black men. It would have been intolerable five centuries ago “to most persons” among the feudal nobility, to hear it asserted that the greatest pleasure or pain of a hundred serfs ought not to give way to the smallest of a nobleman.

And he adds a remarkable statement that commits the defense of utilitarianism to this very question of equal concern for animals:

We are perfectly willing to stake the whole question on this one issue. Granted that any practice causes more pain to animals than it gives pleasure to man; is that practice moral or immoral? And if, exactly in proportion as human beings raise their heads out of the slough of selfishness, they do not with one voice answer “immoral,” let the morality of the principle of utility be forever condemned.

What of Dostoyevsky’s objection, that utilitarianism may tell us to do terrible things, even to torture an innocent child? If we really take the hypothetical scenario seriously, so that unless we torture the child, war and violence will continue (and this will, over the coming centuries, involve the torture and death of many innocent people, both children and adults) then the utilitarian must part company with Alyosha, and accept responsibility for what happens when he or she does not torture the child, as well as what happens when he or she does torture the child. On this view, torturing an innocent child could, conceivably, be the right thing to do, repugnant as that conclusion may be. Of course, it is hard to imagine any real-world scenario in which torturing an innocent child could produce such clearly better consequences, so it is a good thing that we all find the idea of torturing a child to be outrageous.

6. Applications.

From the time of Bentham onwards, utilitarianism has been an active philosophically-grounded social reform movement, not confined to academic discussions, but demanding a wide range of improvements in society in order to reduce suffering and increase happiness, whether for humans or for nonhuman animals. Today contemporary utilitarianism continues to demand change in areas such as aid for the global poor, prison reform, patient choice in death and dying, and the treatment of animals. Today, all these movements are growing in significance. Because utilitarianism insists on giving equal consideration to the interests of future beings, it takes a strong stance on issues that could have very adverse consequences for the future, such as climate change, and on reducing even small risks of human extinction, such as preventing nuclear war, bioterrorism, and tracking asteroids that could collide with our planet and which, if we learn of the threat in time, we may be able to deflect.

Effective altruism is a new movement, broadly utilitarian but not exclusively so, that seeks to encourage more altruism, but also to guide it into causes where it can do the most good. In my recent book *The Most Good You Can Do*, I give many examples of people who have been inspired by broadly utilitarian thinking to become effective altruists. Some people choose to give substantial portions of their income to effective charities. Others change their career plans so as to be able to do more good. Many, to avoid supporting cruel treatment of animals, and also to reduce their greenhouse gas emissions, stop eating meat, or avoid all animal products and become vegan. I even know three people who have donated a kidney to a stranger because they were convinced that the risk to them from losing one kidney is much less than the gain to the stranger who needs a kidney transplant.

In this way utilitarianism continues to be, not only an academically viable normative ethical theory, but a philosophical outlook that moves many people to change their lives in important ways in order to make the world a better place.