



Utilitarianism

That action is best, which procures the greatest happiness for
the greatest numbers.

FRANCIS HUTCHESON, *AN INQUIRY CONCERNING MORAL GOOD AND EVIL*

Suppose you are on an island with a dying millionaire. With his final words, he begs you for one final favor: “I’ve dedicated my whole life to baseball and for fifty years have gotten endless pleasure rooting for the New York Yankees. Now that I am dying, I want to give all my assets, \$2 million, to the Yankees.” Pointing to a box containing money in large bills, he continues: “Would you take this money back to New York and give it to the Yankees’ owner so that he can buy better players?” You agree to carry out his wish, at which point a huge smile of relief and gratitude breaks out on his face as he expires in your arms. After traveling to New York, you see a newspaper advertisement placed by your favorite charity, World Hunger Relief Organization (whose integrity you do not doubt), pleading for \$2 million to be used to save 100,000 people dying of starvation in Africa. Not only will the \$2 million save their lives, but it will also purchase equipment and the kinds of fertilizers necessary to build a sustainable economy. You decide to reconsider your promise to the dying Yankee fan, in light of this advertisement.

What is the right thing to do in this case? Consider some traditional moral principles and see if they help us come to a decision. One principle often given to guide action is “Let your conscience be your guide.” I recall this principle with fondness, for it was the one my father taught me at an early age, and it still echoes in my mind. But does it help here? No, since conscience is primarily a function of upbringing. People’s consciences speak to them in different ways according to how they were brought up. Depending on upbringing, some

people feel no qualms about committing violent acts, whereas others feel the torments of conscience over stepping on a gnat. Suppose your conscience tells you to give the money to the Yankees and my conscience tells me to give the money to the World Hunger Relief Organization. How can we even discuss the matter? If conscience is the end of it, we're left mute.

Another principle urged on us is "Do whatever is most loving"; Jesus in particular set forth the principle "Love your neighbor as yourself." Love is surely a wonderful value. It is a more wholesome attitude than hate, and we should overcome feelings of hate if only for our own psychological health. But is love enough to guide our actions when there is a conflict of interest? "Love is blind," it has been said, "but reason, like marriage, is an eye-opener." Whom should I love in the case of the disbursement of the millionaire's money—the millionaire or the starving people? It's not clear how love alone will settle anything. In fact, it is not obvious that we must always do what is most loving. Should we always treat our enemies in loving ways? Or is it morally permissible to feel hate for those who have purposely and unjustly harmed us, our loved ones, or other innocent people? Should the survivors of Auschwitz love Adolph Hitler? Love alone does not solve difficult moral issues.

A third principle often given to guide our moral actions is the Golden Rule: "Do to others as you would have them do to you." This, too, is a noble rule of thumb, one that works in simple, commonsense situations. But it has problems. First, it cannot be taken literally. Suppose I love to hear loud heavy-metal music. Since I would want you to play it loudly for me, I reason that I should play it loudly for you—even though I know that you hate the stuff. Thus, the rule must be modified: "Do to others as you would have them do to you if you were in their shoes." However, this still has problems. If I were the assassin of Robert Kennedy, I'd want to be released from the penitentiary; but it's not clear that he should be released. If I put myself in the place of a sex-starved individual, I might want to have sex with the next available person; but it's not obvious that I (or anyone else) must comply with that wish. Likewise, the Golden Rule doesn't tell me to whom to give the millionaire's money.

Conscience, love, and the Golden Rule are all worthy rules of thumb to help us through life. They work for most of us, most of the time, in ordinary moral situations. But, in more complicated cases, especially when there are legitimate conflicts of interests, they are limited.

A more promising strategy for solving dilemmas is that of following definite moral rules. Suppose you decided to give the millionaire's money to the Yankees to keep your promise or because to do otherwise would be stealing. The principle you followed would be "Always keep your promise." Principles are

important in life. All learning involves understanding a set of rules; as R. M. Hare says, “Without principles we could not learn anything whatever from our elders.... Every generation would have to start from scratch and teach itself.”¹ If you decided to act on the principle of keeping promises, then you adhered to a type of moral theory called **deontology**. In Chapter 1, we saw that deontological systems maintain that the center of value is the act or kind of act; certain features in the act itself have intrinsic value. For example, a deontologist would see something intrinsically wrong in the very act of lying.

If, on the other hand, you decided to give the money to the World Hunger Relief Organization to save an enormous number of lives and restore economic solvency to the region, you sided with a type of theory called **teleological ethics**. Sometimes, it is referred to as *consequentialist ethics*. We also saw in Chapter 1 that the center of value here is the outcome or consequences of the act. For example, a teleologist would judge whether lying was morally right or wrong by the consequences it produced.

We have already examined one type of teleological ethics: *ethical egoism*, the view that the act that produces the most amount of good for the agent is the right act. Egoism is teleological ethics narrowed to the agent himself or herself. In this chapter, we will consider the dominant version of teleological ethics—*utilitarianism*. Unlike ethical egoism, utilitarianism is a universal teleological system. It calls for the maximization of goodness in society—that is, the greatest goodness for the greatest number—and not merely the good of the agent.

CLASSIC UTILITARIANISM

In our normal lives we use utilitarian reasoning all the time; I might give money to charity when seeing that it would do more good for needy people than it would for me. In time of war, I might join the military and risk dying because I see that society’s needs at that time are greater than my own. One of the earliest examples of utilitarian reasoning is found in Sophocles’ *Antigone* (ca. 440 BCE), in which we find King Creon faced with the tragic task of sacrificing his beloved niece, Antigone, who has violated the law by performing funeral rites over her brother, Polynices. Creon judges that it is necessary to sacrifice one person rather than expose his society to the dangers of rebelliousness—regardless of that person’s innocence or social standing:

And whoever places a friend above the good of his own country, I have no use for him.... I could never stand by silent, watching destruction march against our city, putting safety to rout, nor could I ever make that man a friend of mine who menaces our country. Remember this: our country is our safety.²

As a formal ethical theory, the seeds of utilitarianism were sewn by the ancient Greek philosopher Epicurus (342–270 BCE), who stated that “pleasure is the goal that nature has ordained for us; it is also the standard by which we judge everything good.” According to this view, rightness and wrongness are determined by pleasure or pain that something produces. Epicurus’s theory focused largely on the individual’s personal experience of pleasure and pain, and to that extent he advocated a version of ethical egoism. Nevertheless, Epicurus inspired a series of eighteenth-century philosophers who emphasized the notion of general happiness—that is, the pleasing consequences of actions that impact others and not just the individual. Francis Hutcheson (1694–1746) stated that “that action is best, which procures the greatest happiness for the greatest numbers.” David Hume (1711–1776) introduced the term *utility* to describe the pleasing consequences of actions as they impact people.

The classical expressions of utilitarianism, though, appear in the writings of two English philosophers and social reformers Jeremy Bentham (1748–1832) and John Stuart Mill (1806–1873). They were the nonreligious ancestors of the twentieth-century secular humanists, optimistic about human nature and our ability to solve our problems without recourse to God. Engaged in a struggle for legal as well as moral reform, they were impatient with the rule-bound character of law and morality in eighteenth- and nineteenth-century Great Britain and tried to make the law serve human needs and interests.

Jeremy Bentham

There are two main features of utilitarianism, both of which Bentham articulated: the consequentialist principle (or its teleological aspect) and the utility principle (or its hedonic aspect). The *consequentialist principle* states that the rightness or wrongness of an act is determined by the goodness or badness of the results that flow from it. It is the end, not the means, that counts; the end justifies the means. The *utility*, or *hedonist*, *principle* states that the only thing that is good in itself is some specific type of state (for example, pleasure, happiness, welfare). Hedonistic utilitarianism views pleasure as the sole good and pain as the only evil. To quote Bentham, “Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as what we shall do.”³ An act is right if it either brings about more pleasure than pain or prevents pain, and an act is wrong if it either brings about more pain than pleasure or prevents pleasure from occurring.

Bentham invented a scheme for measuring pleasure and pain that he called the **hedonic calculus**: The quantitative score for any pleasure or pain experience is obtained by summing the seven aspects of a pleasurable or painful experience: its intensity, duration, certainty, nearness, fruitfulness, purity, and extent. Adding up the amounts of pleasure and pain for each possible act and then comparing the scores would enable us to decide which act to perform. With regard to our example of deciding between giving the dying man’s money to the Yankees or to the African famine victims, we would add up the likely pleasures to all involved, for all seven qualities. If we found that giving the money to the

famine victims would cause at least 3 million *hedons* (units of happiness) but that giving the money to the Yankees would cause less than 1,000 hedons, we would have an obligation to give the money to the famine victims.

There is something appealing about Bentham's utilitarianism. It is simple in that there is only one principle to apply: Maximize pleasure and minimize suffering. It is commonsensical in that we think that morality really is about reducing suffering and promoting benevolence. It is scientific: Simply make quantitative measurements and apply the principle impartially, giving no special treatment to ourselves or to anyone else because of race, gender, personal relationship, or religion.

However, Bentham's philosophy may be too simplistic in one way and too complicated in another. It may be too simplistic in that there are values other than pleasure (as we saw in Chapter 6), and it seems too complicated in its artificial hedonic calculus. The calculus is encumbered with too many variables and has problems assigning scores to the variables. For instance, what score do we give a cool drink on a hot day or a warm shower on a cool day? How do we compare a 5-year-old's delight over a new toy with a 30-year-old's delight with a new lover? Can we take your second car from you and give it to Beggar Bob, who does not own a car and would enjoy it more than you? And if it is simply the overall benefits of pleasure that we are measuring, then if Jack or Jill would be "happier" in the Pleasure Machine or the Happiness Machine or on drugs than in the real world, would we not have an obligation to ensure that these conditions obtain? Because of such considerations, Bentham's version of utilitarianism was, even in his own day, referred to as the "pig philosophy" because a pig enjoying his life would constitute a higher moral state than a slightly dissatisfied Socrates.

John Stuart Mill

It was to meet these sorts of objections and save utilitarianism from the charge of being a pig philosophy that Bentham's successor, John Stuart Mill, sought to distinguish happiness from mere sensual pleasure. His version of the theory is often called **eudaimonistic utilitarianism** (from the Greek *eudaimonia*, meaning "happiness"). He defines happiness in terms of certain types of higher-order pleasures or satisfactions such as intellectual, aesthetic, and social enjoyments, as well as in terms of minimal suffering. That is, there are two types of pleasures. The lower, or elementary, include eating, drinking, sexuality, resting, and sensuous titillation. The higher include high culture, scientific knowledge, intellectuality, and creativity. Although the lower pleasures are more intensely gratifying, they also lead to pain when overindulged in. The higher pleasures tend to be more long term, continuous, and gradual.

Mill argued that the higher, or more refined, pleasures are superior to the lower ones: "A being of higher faculties requires more to make him happy, is capable probably of more acute suffering, and certainly accessible to it at more points, than one of an inferior type," but still he is qualitatively better off than the person without these higher faculties. "It is better to be a human being

dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied.”⁴ Humans are the kind of creatures who require more to be truly happy. They want the lower pleasures, but they also want deep friendship, intellectual ability, culture, the ability to create and appreciate art, knowledge, and wisdom.

But one may object, “How do we know that it really is better to have these higher pleasures?” Here, Mill imagines a panel of experts and says that of those who have had a wide experience of pleasures of both kinds almost all give a decided preference to the higher type. Because Mill was an *empiricist*—one who believed that all knowledge and justified belief was based in experience—he had no recourse but to rely on the composite consensus of human history. By this view, people who experience both rock music and classical music will, if they appreciate both, prefer Bach and Beethoven to the Rolling Stones or the Dancing Demons. That is, we generally move up from appreciating simple things (for example, nursery rhymes) to more complex and intricate things (for example, poetry that requires great talent) rather than the other way around.

Mill has been criticized for not giving a better reply—for being an elitist and for unduly favoring the intellectual over the sensual. But he has a point. Don’t we generally agree, if we have experienced both the lower and the higher types of pleasure, that even though a full life would include both, a life with only the former is inadequate for human beings? Isn’t it better to be Socrates dissatisfied than the pig satisfied—and better still to be Socrates satisfied?

The point is not merely that humans wouldn’t be satisfied with what satisfies a pig but that somehow the quality of the higher pleasures is *better*. But what does it mean to speak of better pleasure? The formula he comes up with is this:

Happiness ... [is] not a life of rapture; but moments of such, in an existence made up of few and transitory pains, many and various pleasures, with a decided predominance of the active over the passive, and having as the foundation of the whole, not to expect more from life than it is capable of bestowing.⁵

Mill is clearly pushing the boundaries of the concept of “pleasure” by emphasizing higher qualities such as knowledge, intelligence, freedom, friendship, love, and health. In fact, one might even say that his litmus test for happiness really has little to do with actual pleasure and more to do with a nonhedonic cultivated state of mind.

ACT- AND RULE-UTILITARIANISM

There are two classical types of utilitarianism: act- and rule-utilitarianism. In applying the principle of utility, act-utilitarians, such as Bentham, say that ideally we ought to apply the principle to all of the alternatives open to us at any given moment. We may define act-utilitarianism in this way:

Act-utilitarianism: An act is right if and only if it results in as much good as any available alternative.

One practical problem with act-utilitarianism is that we cannot do the necessary calculations to determine which act is the correct one in each case, for often we must act spontaneously and quickly. So rules of thumb are of practical importance—for example, “In general, don’t lie,” and “Generally, keep your promises.” However, the right act is still that alternative that results in the most utility.

A second problem with act-utilitarianism is that it seems to fly in the face of fundamental intuitions about minimally correct behavior. Consider Richard Brandt’s criticism of act-utilitarianism:

It implies that if you have employed a boy to mow your lawn and he has finished the job and asks for his pay, you should pay him what you promised only if you cannot find a better use for your money. It implies that when you bring home your monthly paycheck you should use it to support your family and yourself only if it cannot be used more effectively to supply the needs of others. It implies that if your father is ill and has no prospect of good in his life, and maintaining him is a drain on the energy and enjoyments of others, then, if you can end his life without provoking any public scandal or setting a bad example, it is your positive duty to take matters into your own hands and bring his life to a close.⁶

The alternative to act-utilitarianism is a view called rule-utilitarianism—elements of which we find in Mill’s theory. Most generally, the position is this:

Rule-utilitarianism: An act is right if and only if it is required by a rule that is itself a member of a set of rules whose acceptance would lead to greater utility for society than any available alternative.

Human beings are rule-following creatures. We learn by adhering to the rules of a given subject, whether it is speaking a language, driving a car, dancing, writing an essay, rock climbing, or cooking. We want to have a set of action-guiding rules by which to live. The act-utilitarian rule, to do the act that maximizes utility, is too general for most purposes. Often, we don’t have time to decide whether lying will produce more utility than truth telling, so we need a more specific rule prescribing truthfulness that passes the test of rational scrutiny. Rule-utilitarianism asserts that the best chance of maximizing utility is by following the *set of rules* most likely to give us our desired results. Because morality is a social and public institution, we need to coordinate our actions with others so that we can have reliable expectations about other people’s behavior.

For the most sophisticated versions of rule-utilitarianism, three levels of rules will guide actions. On the lowest level is a set of utility-maximizing rules of thumb, such as “Don’t lie” and “Don’t cause harm,” that should always be followed unless there is a conflict between them. If these first-order rules conflict, then a second-order set of conflict-resolving rules should be consulted, such as “It’s more important to avoid causing serious harm than to tell the truth.” At the top of the hierarchy is a third-order rule sometimes called the *remainder rule*, which is the principle of act-utilitarianism: When no other rule applies, simply do what your best judgment deems to be the act that will maximize utility.

An illustration of this might be the following: Two of our first-order rules might be “Keep your promises” and “Help those in need when you are not seriously inconvenienced in doing so.” Suppose you promised to meet your teacher at 3 p.m. in his office. On your way there, you come upon an accident victim stranded by the wayside who desperately needs help. It doesn’t take you long to decide to break the appointment with your teacher because it seems obvious in this case that the rule to help others overrides the rule to keep promises. We might say that there is a second-order rule prescribing that the first-order rule of helping people in need when you are not seriously inconvenienced in doing so overrides the rule to keep promises. However, there may be some situation where no obvious rule of thumb applies. Say you have \$50 that you don’t really need now. How should you use this money? Put it into your savings account? Give it to your favorite charity? Use it to throw a party? Here and only here, on the third level, the general act-utility principle applies without any other primary rule; that is, do what in your best judgment will do the most good.

Debates between act- and rule-utilitarians continue today. Kai Nielsen, a staunch act-utilitarian, attacks what he calls *moral conservatism*, which is any normative ethical theory that maintains that there is a privileged moral principle, or cluster of moral principles, prescribing determinate actions that it would always be wrong not to act in accordance with no matter what the consequences. For Nielsen, no rules are sacred; differing situations call forth different actions, and potentially any rule could be overridden.

Nielsen argues further that we are responsible for the consequences of not only the actions that we perform but also the nonactions that we fail to perform. He calls this “negative responsibility.” To illustrate, suppose you are the driver of a trolley car and suddenly discover that your brakes have failed. You are just about to run over five workers on the track ahead of you. However, if you act quickly, you can turn the trolley onto a sidetrack where only one man is working. What should you do? One who makes a strong distinction between *allowing* versus *doing* evil would argue that you should do nothing and merely allow the trolley to kill the five workers. But one who denies that this is an absolute distinction would prescribe that you do something positive to minimize evil. Negative responsibility means that you are going to be responsible for someone’s death in either case. Doing the right thing, the utilitarian urges, means minimizing the amount of evil. So you should actively cause the one death to save the other five lives.⁷ Critics of utilitarianism contend either that negative responsibility is not a strict duty or that it can be worked into other systems besides utilitarianism.

The Strengths of Utilitarianism

Utilitarianism has three very positive features. The first attraction or strength is that it is a single principle, an absolute system with a potential answer for every situation: Do what will promote the most utility! It’s good to have a simple, action-guiding principle that is applicable to every occasion—even if it may be difficult to apply (life’s not simple).

Its second strength is that utilitarianism seems to get to the substance of morality. It is not merely a formal system that simply sets forth broad guidelines for choosing principles but offers no principles—such as the guideline “Do whatever you can universalize.” Rather it has a material core: We should promote human (and possibly animal) flourishing and reduce suffering. The first virtue gives us a clear decision procedure in arriving at our answer about what to do. The second virtue appeals to our sense that morality is made for people and that morality is not so much about rules as about helping people and alleviating the suffering in the world.

As such, utilitarianism seems commonsensical. For instance, it gives us clear and reasonable guidance in dealing with the Kitty Genovese case discussed in Chapter 1: We should call the police or do what is necessary to help her, as long as helping her does not create more disutility than leaving her alone. And, in the case of deciding what to do with the dead millionaire’s \$2 million, something in us says that it is absurd to keep a promise to a dead person when it means allowing hundreds of thousands of famine victims to die. Far more good can be accomplished by helping the needy than by giving the money to the Yankees!

A third strength of utilitarianism is that it is particularly well suited to address the **problem of posterity**—namely, why we should preserve scarce natural resources for the betterment of future generations of humans that do not yet exist. Expressed rhetorically, the question is “Why should I care about posterity; what has posterity ever done for me?” In Chapter 6, we saw that the theory of ethical egoism failed to give us an adequate answer to this problem. That is, the egoist gains nothing by preserving natural resources for future generations that do not yet exist and thus can give no benefit to the egoist. However, utilitarians have one overriding duty: to maximize general happiness. As long as the quality of life of future people promises to be positive, we have an obligation to continue human existence, to produce human beings, and to take whatever actions are necessary to ensure that their quality of life is not only positive but high.

It does not matter that we cannot identify these future people. We may look upon them as mere abstract placeholders for utility and aim at maximizing utility. Derek Parfit explains this using the following utilitarian principle: “It is bad if those who live are worse off than those who might have lived.” He illustrates his principle this way. Suppose our generation has the choice between two energy policies: the “Safe Energy Policy” and the “Risky Energy Policy.”⁸ The Risky Policy promises to be safe for us but is likely to create serious problems for a future generation, say, 200 years from now. The Safe Policy won’t be as beneficial to us but promises to be stable and safe for posterity—those living 200 years from now and beyond. We must choose and we are responsible for the choice that we make. If we choose the Risky Policy, we impose harms on our descendants, even if they don’t now exist. In a sense, we are responsible for the people who will live because our policy decisions will generate different causal chains, resulting in different people being born. But more important, we are responsible for their quality of life because we could have caused human lives to have been better off than they are.

What are our obligations to future people? If utilitarians are correct, we have an obligation to leave posterity to as good a world as we can. This would mean radically simplifying our lifestyles so that we use no more resources than are necessary, keeping as much top soil intact as possible, protecting endangered species, reducing our carbon dioxide emissions, preserving the wilderness, and minimizing our overall deleterious impact on the environment in general while using technology wisely.

CRITICISM OF UTILITARIANISM

Utilitarianism has been around for several centuries, but so too have been its critics, and we need to address a series of standard objections to utilitarianism before we can give it a “philosophically clean bill of health.”

Problems with Formulating Utilitarianism

The first set of problems occurs in the very formulation of utilitarianism: “The greatest happiness for the greatest number.” Notice that we have two “greatest” things in this formula: “happiness” and “number.” Whenever we have two variables, we invite problems of determining which of the variables to rank first when they seem to conflict. To see this point, consider the following example: I am offering a \$1,000 prize to the person who runs the longest distance in the shortest amount of time. Three people participate: Joe runs 5 miles in 31 minutes, John runs 7 miles in 50 minutes, and Jack runs 1 mile in 6 minutes. Who should get the prize? John has fulfilled one part of the requirement (run the longest distance), but Jack has fulfilled the other requirement (run the shortest amount of time).

This is precisely the problem with utilitarianism. On the one hand, we might concern ourselves with spreading happiness around so that the greatest number obtain it (in which case, we should get busy and procreate a larger population). On the other hand, we might be concerned that the greatest possible amount of happiness obtains in society (in which case, we might be tempted to allow some people to become far happier than others, as long as their increase offsets the losers’ diminished happiness). So should we worry more about total happiness or about highest average?

Utilitarians also need to be clear about specifically whose happiness we are talking about: all beings that experience pleasure and pain, or all human beings, or all rational beings. One criterion might exclude mentally deficient human beings, and another might include animals. Finally, utilitarians need to indicate how we measure happiness and make interpersonal comparisons between the happiness of different people. We’ve seen Mill’s efforts to address this problem with his notion of higher pleasures; we’ve also seen the additional complications that his solution creates.

None of the above problems defeat utilitarianism as a workable theory, but they do place a heavy burden on utilitarians to clarify the objectives of their theory.

The Comparative Consequences Objection

Another crucial problem with utilitarianism is that it seems to require a superhuman ability to look into the future and survey a mind-boggling array of consequences of actions. Of course, we normally do not know the long-term consequences of our actions because life is too complex and the consequences go on into the indefinite future. One action causes one state of affairs, which in turn causes another state of affairs, indefinitely, so that calculation becomes impossible. Recall the nursery rhyme:

For want of a nail, the shoe was lost;
 For want of a shoe, the horse was lost;
 For want of a horse, the rider was lost;
 For want of a rider, the battle was lost;
 For want of a battle, the kingdom was lost;
 And all for the want of a horseshoe nail.

Poor, unfortunate blacksmith; what utilitarian guilt he must bear all the rest of his days!

But it is ridiculous to blame the loss of one's kingdom on the poor, unsuccessful blacksmith, and utilitarians are not so foolish as to hold him responsible for the bad situation. Instead, following C. I. Lewis, utilitarians distinguish two kinds of consequences: (1) actual consequences of an act and (2) consequences that could reasonably have been expected to occur.⁹ Based on these two kinds of consequences, there are two corresponding right actions. An act is *absolutely* right if it has the best actual consequences (as per consequence 1). An act is *objectively* right if it is reasonable to expect that it will have the best consequences (as per consequence 2).

Only *objective rightness*, that based on reasonable expectations, is central here. *Actual rightness*, based on actual consequences, is irrelevant because this can only be determined after an action is performed and we sit back and watch the series of actual consequences unfold. But when an agent is trying to determine in advance how to act, the most that she can do is to use the best information available and do what a reasonable person would expect to produce the best overall results. Suppose, for example, that while Hitler's grandmother was carrying little Adolph up the stairs to her home, she slipped and had to choose between either dropping infant Adolph and allowing him to be fatally injured or breaking her arm. According to the formula just given, it would have been *absolutely* right for her to let him be killed because history would have turned out better. But, it would not have been within her power to know that. She did what any reasonable person would do—she saved the baby's life at the risk of injury to herself. She did what was *objectively* right. The utilitarian theory holds that by generally doing what reason judges to be the best act based on likely consequences, we will, in general, actually promote the best consequences.

The Consistency Objection to Rule-Utilitarianism

An often-debated question about rule-utilitarianism is whether, when pushed to its logical limits, it must either become a deontological system or transform itself into act-utilitarianism. As such, it is an inconsistent theory that offers no truly independent standard for making moral judgments. Briefly, the argument goes like this: Imagine that following the set of general rules of a rule-utilitarian system yields 100 hedons (positive utility units). We could always find a case where breaking the general rule would result in additional hedons without decreasing the sum of the whole. So, for example, we could imagine a situation in which breaking the general rule “Never lie” to spare someone’s feelings would create more utility (for example, 102 hedons) than keeping the rule would. It would seem that we could always improve on any version of rule-utilitarianism by breaking the set of rules whenever we judge that by doing so we could produce even more utility than by following the set.

To illustrate more fully, consider this example. Suppose a disreputable former convict named Charley has been convicted of a serious crime and sentenced to a severe punishment. You, the presiding judge, have just obtained fresh evidence that if brought into court would exonerate Charley of the crime. But you also have evidence, not admissible in court, that Charley is guilty of an equally heinous crime for which he has not been indicted. The evidence suggests that Charley is a dangerous man who should not be on the streets of our city. What should you do? An act-utilitarian would no doubt suppress the new evidence in favor of protecting the public from a criminal. A rule-utilitarian has a tougher time making the decision. On the one hand, he has the rule “Do not permit innocent people to suffer for crimes they didn’t commit.” On the other hand, he has the rule “Protect the public from unnecessary harm.” The rule-utilitarian may decide the matter by using the *remainder principle*, which yields the same result as that of the act-utilitarian. This seems, however, to give us a counterintuitive result. Why not just be an act-utilitarian and forgo the middle steps if that is what we are destined to reach anyway?

There may be other ways for the rule-utilitarian to approach this. He or she may opt for a different remainder principle, one that appeals to our deepest intuitions: “Whenever two rules conflict, choose the one that fits your deepest moral intuition.” Thus, the judge may very well decide to reveal the evidence exonerating Charley, holding to the rule not to allow people to suffer for crimes for which there is insufficient evidence to convict them. The rule-utilitarian argues that, in the long run, a *rule* that protects such legally innocent but morally culpable, people will produce more utility than following an act-utilitarian principle. If we accept the second intuitionist version of the remainder principle, we may be accused of being deontological intuitionists and not utilitarians at all.

How might we respond to this criticism of inconsistency? It may be more accurate to see moral philosophy as complex and multidimensional so that both striving for the goal of utility and the method of consulting our intuitions are part of moral deliberation and action. Thus, even if rule-utilitarianism involves consulting moral intuitions, both of these elements may be intertwined and

equally legitimate parts of moral reasoning. What at first appears to be a problem of consistency is really just an indicator of the multilayered nature of morality.

The No-Rest Objection

According to utilitarianism, one should always do that act that promises to promote the most utility. But there is usually an infinite set of possible acts to choose from, and even if I can be excused from considering all of them, I can be fairly sure that there is often a preferable act that I could be doing. For example, when I am about to go to the cinema with a friend, I should ask myself if helping the homeless in my community wouldn't promote more utility. When I am about to go to sleep, I should ask myself whether I could at that moment be doing something to help save the ozone layer. And, why not simply give all my assets (beyond what is absolutely necessary to keep me alive) to the poor to promote utility? Following utilitarianism, I should get little or no rest, and, certainly, I have no right to enjoy life when by sacrificing I can make others happier. Peter Singer actually advocates an act-utilitarian position similar to this. According to Singer, middle-class people have a duty to contribute to poor people (especially in undeveloped countries) more than one-third of their income, and all of us have a duty to contribute every penny above \$30,000 we possess until we are only marginally better off than the worst-off people on earth.

The problem with approaches like Singer's is that this makes morality too demanding, creates a disincentive to work, and fails to account for different levels of obligation. Thus, utilitarianism must be a false doctrine. But rule-utilitarians have a response to this no-rest objection: A rule prescribing rest and entertainment is actually the kind of rule that would have a place in a utility-maximizing set of rules. The agent should aim at maximizing his or her own happiness as well as other people's happiness. For the same reason, it is best not to worry much about the needs of those not in our primary circle. Although we should be concerned about the needs of poor people, it actually would promote disutility for the average person to become preoccupied with these concerns. Singer represents a radical act-utilitarian position that fails to give adequate attention to the rules that promote human flourishing, such as the right to own property, educate one's children, and improve one's quality of life, all of which probably costs more than \$30,000 per year in many parts of North America. However, the utilitarian would remind us, we can surely do a lot more for suffering humanity than we now are doing—especially if we join together and act cooperatively. And we can simplify our lives, cutting back on unnecessary consumption, while improving our overall quality.

The Publicity Objection

It is usually thought that moral principles must be known to all so that all may freely obey the principles. But utilitarians usually hesitate to recommend that everyone act as a utilitarian, especially an act-utilitarian, because it takes a great deal of deliberation to work out the likely consequences of alternative courses of

action. It would be better if most people acted simply as deontologists.¹⁰ Thus, utilitarianism seems to contradict our requirement of publicity.

There are two responses to this objection. First, at best this objection only works against act-utilitarianism, which at least in theory advocates sitting down and calculating the good and bad consequences of each action that we plan to perform. Rule-utilitarianism, by contrast, does not focus on the consequences of particular actions but on the set of rules that are likely to bring about the most good. These rules indeed are publicized by rule-utilitarians.

A second response is one that act-utilitarians themselves might offer: The objection shows a bias only toward publicity (or even democracy). It may well be that publicity is only a rule of thumb to be overridden whenever there is good reason to believe that we can obtain more utility by not publicizing act-utilitarian ideas.

However, this response places an unacceptably low value on the benefits of publicity. Since we need to coordinate our actions with other people, moral rules must be publicly announced, typically through legal statutes. I may profit from cutting across the grass to save a few minutes in getting to class, but I also value a beautiful green lawn. We need public rules to ensure the healthy state of the lawn. So we agree on a rule to prohibit walking on the grass—even when it may have a utility function. There are many activities that may bring about individual utility advancement or even communal good, which if done regularly would be disastrous, such as cutting down trees to build houses or make newspapers or paper for books, valuable as it is. So we regulate the lumber industry so that every tree cut down is replaced with a new one and large forests are kept intact. So moral rules must be publicly advertised, often made into laws, and enforced. In short, while the publicity objection does not affect rule-utilitarianism, it appears to be a serious obstacle to act-utilitarianism.

The Relativism Objection

Sometimes people accuse rule-utilitarianism of being relativistic because it seems to endorse different rules in different societies. In one society, it may uphold polygamy, whereas in our society it defends monogamy. In a desert society, it upholds a rule “Don’t waste water,” whereas in a community where water is plentiful no such rule exists. But this is not really conventional relativism because the rule is not made valid by the community’s choosing it but by the actual situation. In the first case, it is made valid by an imbalance in the ratio of women to men and, in the second case, by the environmental factors concerning the availability of water. Situationalism is different from relativism and consistent with objectivism because it really has to do with the application of moral principles—in this case, the utility principle.

But there is a more serious worry about rule-utilitarianism’s tendency toward relativism—namely, that it might become so plastic that it justifies any moral rule. Asked why we support benevolence as a moral rule, it seems too easy to respond, “Well, this principle will likely contribute to the greater utility in the long run.” The fear is that the act-utilitarian could give the same answer

to rules that we consider malevolent, such as torture. Shifting conceptions of general happiness will generate shifting moral rules.

How might the rule-utilitarian respond to this? David Hume, an early defender of utilitarian moral reasoning, argued that human nature forces consistency in our moral assessments. Specifically, he argues, there are “universal principles of the human frame” that regulate what we find to be agreeable or disagreeable in moral matters. Benevolence, for example, is one such type of conduct that we naturally find agreeable.¹¹ Following Hume’s lead, the rule-utilitarian might ground the key components of happiness in our common human psychological makeup rather than the result of fluctuating personal whims. This would give utilitarianism a more objective foundation and thus make it less susceptible to the charge of relativism.

CRITICISM OF THE ENDS JUSTIFYING IMMORAL MEANS

Chief among the criticisms of utilitarianism is that utilitarian ends might justify immoral means. There are many dastardly things that we can do in the name of maximizing general happiness: deceit, torture, slavery, even killing off ethnic minorities. As long as the larger populace benefits, these actions might be justified. The general problem can be laid out in this argument:

- (1) If a moral theory justifies actions that we universally deem impermissible, then that moral theory must be rejected.
- (2) Utilitarianism justifies actions that we universally deem impermissible.
- (3) Therefore, utilitarianism must be rejected.

Let’s look at several versions of this argument.

The Lying Objection

William D. Ross has argued that utilitarianism is to be rejected because it leads to the counterintuitive endorsement of lying when it serves the greater good. Consider two acts, A and B, that will both result in 100 hedons (units of pleasure of utility). The only difference is that A involves telling a lie and B involves telling the truth. The utilitarian must maintain that the two acts are of equal value. But this seems implausible; truth seems to be an intrinsically good thing.

Similarly, in Arthur Koestler’s *Darkness at Noon*, we find this discussion of Communist philosophy in the former Soviet Union:

History has taught us that often lies serve her better than the truth; for man is sluggish and has to be led through the desert for forty years before each step in his development. And he has to be driven through the desert with threats and promises, by imaginary terrors and imaginary consolations, so that he should not sit down prematurely to rest and divert himself by worshipping golden calves.¹²

According to this interpretation, orthodox Soviet communism justified its lies through utilitarian ideas. Something in us revolts at this kind of value system. Truth is sacred and must not be sacrificed on the altar of expediency.

In response to this objection, utilitarians might agree that there is something counterintuitive in the calculus of equating an act of lying with one of honesty; but, they argue, we must be ready to change our culturally induced moral biases. What is so important about truth telling or so bad about lying? If it turned out that lying really promoted human welfare, we'd have to accept it. But that's not likely. Our happiness is tied up with a need for reliable information (that is, truth) on how to achieve our ends, so truthfulness will be a member of the rule-utility's set. But where lying will clearly promote utility without undermining the general adherence to the rule, we simply ought to lie. Don't we already accept lying to a gangster or telling white lies to spare people's feelings?

The Integrity Objection

Bernard Williams argues that utilitarianism violates personal integrity by commanding that we violate our most central and deeply held principles. He illustrates this with the following example:

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 sweat-stained 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 inhabitants who, after recent acts of protest against the government, are just about to be killed to remind other possible protesters of the advantages of not protesting. However, since Jim is an honored 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 setup that nothing of that kind is going to work: any attempt of that sort of thing will mean that all the Indians will be killed, and himself. The men against the wall, the other villagers, understand the situation, and are obviously begging him to accept. What should he do?¹³

Williams asks rhetorically,

How can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which

he has built his life, just because someone else's projects have so structured the causal scene that *that* is how the utilitarian sum comes out?¹⁴

In response to this criticism, the utilitarian can argue that integrity is not an absolute that must be adhered to at all costs. Some alienation may be necessary for the moral life, and the utilitarian can take this into account in devising strategies of action. Even when it is required that we sacrifice our lives or limit our freedom for others, we may have to limit or sacrifice something of what Williams calls our integrity. We may have to do the “lesser of evils” in many cases. If the utilitarian doctrine of negative responsibility is correct, we need to realize that we are responsible for the evil that we knowingly allow, as well as for the evil we commit.

The Justice Objection

With both of the above problems, the utilitarian response was that we should reconsider whether truth telling and personal integrity are values that should never be compromised. The situation is intensified, though, when we consider standards of justice that most of us think should never be dispensed with. Let's look at two examples, each of which highlights a different aspect of justice.

First, imagine that a rape and murder is committed in a racially volatile community. As the sheriff of the town, you have spent a lifetime working for racial harmony. Now, just when your goal is being realized, this incident occurs. The crime is thought to be racially motivated, and a riot is about to break out that will very likely result in the death of several people and create long-lasting racial antagonism. You see that you could frame a tramp for the crime so that a trial will find him guilty and he will be executed. There is every reason to believe that a speedy trial and execution will head off the riot and save community harmony. Only you (and the real criminal, who will keep quiet about it) will know that an innocent man has been tried and executed. What is the morally right thing to do? The utilitarian seems committed to framing the tramp, but many would find this appalling.

As a second illustration, imagine that you are a utilitarian physician who has five patients under your care. One needs a heart transplant, one needs two lungs, one needs a liver, and the last two each need a kidney. Now into your office comes a healthy bachelor needing an immunization. You judge that he would make a perfect sacrifice for your five patients. Through a utility-calculus, you determine that, without a doubt, you could do the most good by injecting the healthy man with a fatal drug and then using his organs to save your five other patients.¹⁵

These careless views of justice offend us. The very fact that utilitarians even consider such actions—that they would misuse the legal system or the medical system to carry out their schemes—seems frightening. It reminds us of the medieval Roman Catholic bishop's justification for heresy hunts and inquisitions and religious wars:

When the existence of the Church is threatened, she is released from the commandments of morality. With unity as the end, the use of every

means is sanctified, even cunning, treachery, violence, simony, prison, death. For all order is for the sake of the community, and the individual must be sacrificed to the common good.¹⁶

Similarly, Koestler argues that this logic was used by the Communists in the Soviet Union to destroy innocent people whenever it seemed to the Communist leaders that torture and false confessions served the good of the state because “you can’t make an omelet without breaking eggs.”

How can the utilitarian respond to this? It won’t work this time to simply state that justice is not an absolute value that can be overridden for the good of the whole society. The sophisticated rule-utilitarian insists it makes good sense to have a principle of justice to which we generally adhere. That is, general happiness is best served when we adopt the value of justice. Justice should not be overridden by current utility concerns because human rights themselves are outcomes of utility consideration and should not be lightly violated. That is, because we tend subconsciously to favor our own interests and biases, we institute the principle of rights to protect ourselves and others from capricious and biased acts that would in the long run have great disutility. Thus, we must not undermine institutional rights too easily. Thus, from an initial rule-utilitarian assessment, the sheriff should not frame the innocent tramp, and the doctor should not harvest organs from the bachelor.

However, the utilitarian cannot exclude the possibility of sacrificing innocent people for the greater good of humanity. Wouldn’t we all agree that it would be right to sacrifice one innocent person to prevent an enormous evil? Suppose, for example, a maniac is about to set off a nuclear bomb that will destroy New York City. He is scheduled to detonate the bomb in one hour. His psychiatrist knows the lunatic well and assures us that there is one way to stop him—torture his 10-year-old daughter and televise it. Suppose for the sake of the argument that there is no way to simulate the torture. Would you not consider torturing the child in this situation? As the rule-utilitarian would see it, we have two moral rules that are in conflict: the rule to prevent widespread harm and the rule against torture. To resolve this conflict, the rule-utilitarian might appeal to this second-level conflict-resolving rule: We may sacrifice an innocent person to prevent a significantly greater social harm. Or, if no conflict-resolving rule is available, the rule-utilitarian can appeal to this third-level remainder rule: When no other rule applies, simply do what your best judgment deems to be the act that will maximize utility. Using this remainder rule, the rule-utilitarian could justify torturing the girl.

Thus, in such cases, it might be right to sacrifice one innocent person to save a city or prevent some wide-scale disaster. In these cases, the rule-utilitarian’s approach to justice is in fact the same as the above approach to lying and compromising one’s integrity: Justice is just one more lower-order principle within utilitarianism. The problem, clearly, is determining which kinds of wide-scale disasters warrant sacrificing innocent lives. This question invariably comes up in wartime: In every bombing raid, especially in the dropping of the atomic bomb on Hiroshima and Nagasaki, the noncombatant–combatant distinction is overridden. Innocent civilian lives are sacrificed with the prospect of

ending the war. We seem to be making this judgment call in our decision to drive automobiles and trucks even though we are fairly certain the practice will result in the death of thousands of innocent people each year. Judgment calls like these highlight utilitarianism's difficulty in handling issues of justice.

CONCLUSION

We've seen that multilevel rule-utilitarianism satisfies the purposes of ethics, gives a clear decision procedure for moral conduct, and focuses on helping people and reducing suffering in the world. It also offers a compelling solution to the problem of posterity. Further, rule-utilitarianism has responses to all the criticisms directed toward it. Whether the responses are adequate is another story. Perhaps it would be better to hold off making a final judgment about utilitarianism until considering the next two chapters, in which two other types of ethical theory are discussed.

NOTES

1. R. M. Hare, *The Language of Morals* (Oxford: Oxford University Press, 1952), p. 60.
2. Sophocles, *Antigone*, trans. Robert Eagles (Penguin Classics, 1982), lines 204–214.
3. Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (1789), Ch. 1.
4. J. S. Mill, *Utilitarianism* (1861), Ch. 2.
5. *Ibid.*
6. Richard Brandt, "Towards a Credible Form of Utilitarianism," in *Morality and the Language of Conduct*, ed. H. Castaneda and G. Naknikian (Wayne State University Press, 1963), pp. 109–110.
7. Kai Nielsen, "Against Moral Conservatism," *Ethics* 82 (1972): 219–231.
8. Derek Parfit, "Energy Policy and the Further Future: The Identity Problem," in *Energy and the Future*, ed. D. MacLean and P. Brown (Rowman & Littlefield, 1983).
9. See Anthony Quinton, *Utilitarian Ethics* (Macmillan, 1973), pp. 49–50. Lewis and Quinton add a third type of consequence, namely, intended ones: An act is *subjectively* right if its agent intends or actually expects it to have the best consequences.
10. The famous nineteenth-century utilitarian Henry Sidgwick, in his *The Methods of Ethics* (Oxford, 1974), p. 483, argues that utilitarians should keep their views a secret for the good of society.
11. David Hume, *An Enquiry Concerning the Principles of Morals* (1751), Sec. 9, Pt. 1.
12. Arthur Koestler, *Darkness at Noon* (Macmillan, 1941), p. 80.
13. Bernard Williams, "A Critique of Utilitarianism," in *Utilitarianism: For and Against* (Smart & Williams), pp. 98–99.
14. *Ibid.*