



and large do not posit goals that must be followed but instead set side constraints on one's actions. "Don't commit murder," "Don't tell lies," "Keep your promises" are examples of such rules.

An example that illustrates the conflict: Suppose one is a surgeon. Six patients enter one's office at once. Five of them unfortunately are gravely ill. Each of the five must receive an organ transplant very soon or he will die. One needs a heart, one needs a kidney, etc.--five different organs are needed. But fortunately, the sixth man who wandered into the office has all the healthy organs needed. The surgeon faces a choice between killing the healthy patient in order to save the five, and refusing to kill the healthy patient, thereby letting the five die. (In the example these are the only possible choices--it won't work to wait till one of the diseased patients dies, then cut up his corpse and use his organs to save the four who are threatened. By the time the first threatened person dies, his organs will be useless for transplant purposes.) What should the surgeon do? The common-sense moral rule "Don't murder" tells her that she should refrain from cutting up the one even in order to save the five. But it appears that by act-utilitarian calculation five deaths are worse than one death, so in this situation the doctor ought to kill the one in order to save the five. Common-sense morality and utilitarianism appear to be in sharp disagreement.

### **Utility: Rival Views.**

Utility is a name for whatever makes someone's life go best for that very person.. To put it another way: "utility" refers to whatever is intrinsically valuable in a human life. What is intrinsically valuable is valuable for its own sake, or in itself. What is extrinsically valuable is valuable in that it contributes to achieving some further value. (For example, having money is valuable for getting a can of Coke from the soft drink machine, and getting the can is valuable for getting Coke in one's mouth, but the enjoyment of the taste of the Coke in one's mouth is valuable for itself, not for anything further it helps to bring about.) We can distinguish importantly different views about what utility is. Start with J. S. Mill's official view, hedonism.

**Hedonism** on a narrow construal says that utility just is pleasure and avoidance of pain. The greater the overall total of pleasure minus pain in a life, the better the life. A broader construal of hedonism identifies utility with desired experience Another construal, suggested by some of Mill's remarks in chapter 2 of *Utilitarianism* identifies utility with rationally desired pleasure--pleasure that would be desired by experienced and competent judges. The root idea of these hedonistic views is that when we are considering whether a person's life went well or badly on the whole, nothing counts except the experiences of that person. What matters is how life feels to a person "from the inside," as she lives it. This is the view that Robert Nozick aims to challenge with his "experience machine" discussion.

Welfarism or the **Desire Satisfaction View** holds that utility consists entirely in satisfaction of desires. More exactly, if we distinguish basic from instrumental desires (a desire is basic if what is desired is desired for its own sake; instrumental desires are desires for things as means to further goals), welfarism holds that utility consists in the satisfaction of basic desires. The more it is the case that a person gets what he wants for its own sake, the better his life has gone. A variant of welfarism identifies utility with satisfaction of rational or fully informed desires. A person's rational desires, let us say, are the desires she would have if she were to be in possession of full pertinent information and making no cognitive errors. Example: Suppose my wife's overriding desire is to build a monument to her husband's virtue (a huge statue in the back yard, say). Building this monument is her ultimate goal in life. But suppose she formed this desire only as a result of reasoning incorrectly to the conclusion that I am extraordinarily virtuous, which I am not, and that if she figured out the truth this overriding desire would extinguish itself. In this case my wife's desire is not rational even if she now thinks otherwise. The closer a person's actual desires are to her fully informed desires, the more it is the case that satisfying her actual desires enhances her welfare.

What information is pertinent to the quality of one's desires? One proposal is that any information that would cause my basic desires to change is pertinent (relevant) information. Information that is such that my learning it would not cause my basic desires to shift is not pertinent information.

As formulated, the fully informed desire satisfaction view seems vulnerable to a problem. Suppose I want to desire to learn quantum physics-not as a means to any further goal, just for the sake of gaining this knowledge. But if I were fully informed, I would already know quantum physics, and then it would be pointless to desire to learn it. Moreover, information about quantum physics is pertinent to my basic desire to learn quantum physics, because having this information would presumably cause my basic desires to shift. Sometimes the fully informed desire satisfaction view is slightly altered to deal with this problem. The proposal is that my life goes well to the extent that

my actual basic desires (a) conform to the desires my ideal adviser would want me to have and (b) are satisfied. My ideal adviser would be a person exactly like me, with my psychology and traits, except that he is fully informed and making no cognitive errors. My ideal adviser is specified as being someone who is sympathetic to me. What the ideal adviser would want me to want fixes the desires, satisfaction of which constitutes my life going well for me.

According to desire satisfaction views, a person's life goes well to the extent that she gets what she wants for its own sake over the course of her life, with her wants weighted according to her own rating of their importance. One problem with such views is that not all of a person's desires intuitively seem to be such that their satisfaction advances the person's welfare. Consider the example of meeting a stranger on a train, having a conversation with her, then separating and meeting no more. If one forms the desire that the stranger's life go well, the satisfaction of this desire does not seem to make one's own life go better. Notice also that one might desire above all for its own sake that one's own life should go badly (suppose one feels guilty and wants to punish oneself). The satisfaction of this desire that one have a bad life surely does not render it the case that one has a good life or contribute to one's good. So not all basic desires, if satisfied, increase one's welfare. Which ones do? There is a threat of circularity here. If one says, the satisfaction of the desires concerned with one's own welfare advance one's welfare, then one must have a prior idea of one's welfare (what is good for oneself), and if one has that, one does not need a further account of what constitutes welfare such as the desire satisfaction view claims to offer.

We should distinguish a life that is good in the sense of fine or admirable or choiceworthy and a life that is good in the sense of good for the one who lives it. A life in which the person wants to advance some noble cause and sacrifices her own good to the cause might be an admirable or choiceworthy life but not one that is good for the agent. Sometimes this line is difficult to draw. A parent may rightly feel that to a degree, her own life goes better when her child's life goes better, especially when the child's life goes better via the agency of the parent. A person who takes on a project such as saving the whales and contributes significantly to it may find her own good bound up with the success of the project.

Leaving these objections to the side, one might still wonder whether it is plausible to identify one's good (what is good for one) or welfare even with rational desire satisfaction. Consider an ideally coherent anorexic. She prefers conforming to her notion of a thin body ideal even at the cost of death at a young age by starvation. We might imagine this person as knowing all the relevant facts and affirming the extreme thin body ideal for herself without making any cognitive errors (she does not add up two and two and get the answer five). But one might hesitate to identify the good for this person with the satisfaction of her most important basic desires even though her desires are not based on factual illusion or confused reasoning. This hesitation indicates an inclination to adopt an Objective List theory of the good.

The desire satisfaction conception of utility or welfare or good is subjectivist. A *subjectivist* conception of utility is any conception that holds that what is good for a person cannot be determined independently of that very person's tastes, desires, or values (perhaps as they would be if corrected by well-informed rational scrutiny). The subjectivist holds that what is good for a person is relative to that very person's perspective.

Counterposed to all subjectivist views about the nature of utility is the doctrine known as the **Objective List Theory**. It holds that utility consists in the achievement of objectively valuable goods. On this view, whether some putative good is really good for a person can in principle be determined independently of that very person's tastes, values, or desires regarding that thing. The Objective List theorist thinks that we can know that certain things make someone's life go better. We arrive at a list of such goods. The question of how intrinsically valuable a person's life was can then be settled, according to this view, by determining to what extent the person over the course of her life achieves the goods on the objective list. For example, if we think that friendship, love, athletic prowess, religious ecstasy, artistic creativity, and intellectual achievement are the objectively valuable goods, we check to what extent a person's life scores high on these various dimensions and sum the total to determine the person's "human well-being score." Stated this baldly, the view may sound silly, but if we think about it, it does seem that many private and public policy judgments that people make do presuppose our ability to make and verify at least some judgments of this kind.

(Hedonism is then an instance of an Objective List view. Hedonism holds (a) that an objective list of goods determines what is good for people and (b) that the objective list contains one entry, namely, pleasure. You might be of the opinion that pleasures are fleeting and inherently worthless. No matter, says the hedonist. According to

hedonism, pleasure alone is intrinsically good for a person, whatever might be her own tastes, values, or desires. To ensure that the categories are distinct, we might stipulate that according to an Objective List conception of good, (a) the good life for a person consists in getting items on the list, and (b) the list contains more than one entry.)

Within the set of Objective List views there is an important family of views called "perfectionist." The perfectionist holds that human good, the good life for persons, consists (almost entirely) in the perfection of human nature--the development and exercise of capacities that are truly worthy. Some human capacities are of marginal worth, at best, such as the capacity to fart and the capacity to make other people uncomfortable. But some capacities qualify as excellences, as genuinely worthy. The maximal development and exercise of these excellences counts as human perfection. In our course readings, Robert Adams defends a sort of hybrid view that combines perfectionism and hedonism. He identifies human good with enjoyment of the excellent. Or at least, he considers this position even if he does not clearly affirm it.

Other hybrid views are possible. For example, one might hold that human good is desiring what is objectively valuable (entries on the Objective List) and getting what you desire over the course of your life. Or one might hold that good is getting what is at once objectively valuable, desired by the agent, and experienced as pleasureable. Or one might hold a hybrid or mixed view combining hedonism and some version of the desire satisfaction theory.

There are two obvious difficulties any Objective List view must face. Any hybrid view that includes an Objective List component will confront these difficulties. One difficulty stems from **disagreement** as to what items belong on the Objective List. Suppose two people disagree about this. How are we to settle the dispute by showing that one or both is mistaken? Another difficulty is **alienation**. We might think that what is good for a person must have the power to attract and motivate her at least under favorable conditions. But it seems that I might acknowledge that playing chess belongs on the Objective List while feeling no motivation at all to include chess playing in my plan of life.

The Objective List theory might seem to be faced with another problem, rigidity. It might be thought that the Objective List theorist must hold there is one List for all persons. But I doubt this is so. There are at least three possible versions of the Objective List Theory: (1) there is one list for all persons; (2) there are different types of persons differing in their natures and a distinct and possibly different Objective List for each type, so that what is good for a person depends on her type; (3) each person is unique, has a unique individual nature, so there is a distinct and possibly different Objective List for each person. The advocate of version (3) will still maintain that a person can be mistaken about what is ultimately good for her (what the entries are on her Objective List).

Various examples discussed in class such as the experience machine, the ideally coherent anorexic, the burn victim example, and the deluded professor example can be used to clarify what is involved in agreeing or disagreeing with the utility theories sketched above.

### **Mill on Happiness.**

Mill calls himself a utilitarian. In the terminology introduced above, Mill appears to espouse something close to act-utilitarianism plus a hedonistic theory of utility. At least, in paragraph 2 of chapter 2 of Utilitarianism, Mill writes that the utilitarian is one who "holds that acts are right in proportion as they tend to promote happiness; wrong as they tend to promote the reverse of happiness," and this sounds somewhat like act-utilitarianism except that the latter view holds that of all the acts one could do at a time, the one right choice is the one that maximizes utility and all other choices are wrong. Mill's "righter/wronger" test, in contrast, is scalar: a choice is not judged right or wrong simply, but more or less right or more or less wrong depending on whether or not it tends to produce happiness or unhappiness. Another complication is that Mill does not say that acts are right or wrong depending on the degree to which they produce happiness or unhappiness. He says rather that acts are right or wrong to the degree that they tend to produce happiness or unhappiness. Mill might mean that an act tends to produce happiness if it reasonably expected by the agent at the time of choice to produce happiness (whether it actually does so or not). Mill might alternatively mean that types or kinds of acts, not particular acts, are judged more or less right or wrong by the principle. Types of acts tend to produce happiness or unhappiness. (Does Mill mean to assert that an act is more or less right or wrong depending on whether it is of a type that generally tends to produce more or less happiness or unhappiness, even if on this particular occasion the choice of the act would produce the reverse of what the class of acts to which it belongs tends to produce?) Maybe Mill means: The morally right act, the one one morally ought

to do, is one that, among the available alternatives, would produce an outcome no worse than anything else one might have done instead. Outcomes are better or worse, depending on the net amount of happiness (pleasure minus pain) they contain. The larger the shortfall between the outcome of the act one actually does and the best one might have done instead, the "wronger" the act.

Mill's version of utilitarianism identifies utility with happiness and happiness with "pleasure and the absence of pain." Mill does not explicitly consider whether one might hold that happiness or utility is something other than pleasure. He seems somewhat uncomfortable with the idea that happiness just consists in pleasure, however. In paragraphs 3-9 of chapter 2 he proceeds to defend himself against the objection that the utilitarian who recommends maximizing pleasure in human life is recommending that people should above all devote themselves to satisfying their lower pleasures, the pleasures of sensation such as drinking beer, eating, and engaging in other simple bodily functions. It is assumed by the objector and not denied by Mill that these lower pleasures are greater in intensity than the higher pleasures of the intellect. One possible response to the objection (see paragraphs 4 and 9) is to say that indulging in the lower pleasures usually does no good to anybody except the agent himself. In contrast, indulging in the higher pleasures often benefits others. The person who gets drunk gives pleasure only to himself (and possibly a crony), whereas the person who writes a great novel or discovers a polio vaccine confers vast pleasures on other people. So on grounds of greater fruitfulness we ought to promote the higher pleasures. Also, relentless pursuit of intense sensations tends to wear out the body and thus to produce displeasure in the long run. Mill is not satisfied with these responses. He wants to argue that the higher pleasures are superior as pleasures to the lower pleasures—superior in themselves and without consideration of their further consequences or moral reasons for preferring one sort of pleasure to another. The argument he develops is complex and not obviously convincing. He distinguishes between the quality and the quantity of a pleasure, the test for both being the choice of informed experts. (Or is Mill saying rather that the choices of the informed experts are evidence that particular pleasures are high-quality or low-quality but do not make it the case that they are one or the other? In football, scoring more points than the opposing team constitutes winning the game. There is no conceptual room for the thought that the Chargers might have lost the game even though they scored more points. In contrast, the verdict of a jury in a criminal trial does not constitute guilt or innocence. A guilty accused individual might be found innocent and an innocent accused individual might be found guilty by the jury in a criminal trial.) He distinguishes between happiness and contentment and claims that the fan of the higher pleasures may be less contented, but (other things equal) will be happier than the fan of the lower pleasures. He also discusses what he calls a "sense of dignity," which is supposed to explain why the wise person, the instructed person, and the sensitive person would not trade places with their opposite counterparts, "even though they should be persuaded that the fool, the dunce, or the rascal is better satisfied with his lot than they are with theirs." I invite you to investigate to what extent these various arguments bolster Mill's case against the initial objection. Another question: is Mill's final position a version of hedonism or instead a hybrid view, roughly that nothing is intrinsically good for a person unless it is pleasureable, but the superior, qualitatively better pleasures are those that involve the development and exercise of complex rational capacities (these are the ones the competent judges will prefer)?

### **The Place of Rules in Utilitarianism: Mill's View**

The last ten or so pages of chapter 2 of *Utilitarianism* consider what should be the place of ordinary moral rules in utilitarian thinking. One issue here is that utilitarianism might seem a doctrine fit for angels, not for humans. We humans are (1) not perfectly accurate and fast reasoners, (2) not perfectly well informed about the empirical facts that bear on the proper choice of action and policy, and (3) not perfectly impartial but rather tend to be partial to themselves and to those who are near and dear. Given these weaknesses, we need strict rules to guide our conduct. Telling us we should just "maximize utility" is a recipe for disaster. Given this guidance, people will tend too often to break useful moral rules because they are either incapable of calculating how utility might best be maximized or they will make the calculation with a thumb on the scale tilting the outcome of calculation in favor of their own personal interests. Mill denies that anything that is true in this line of thought constitutes a valid objection against utilitarianism. Some of the points he makes in this connection are these:

1. Utilitarianism is a test of right and wrong action. It does not claim that we should be motivated to maximize utility in all that we do. Institutions and social rules should be arranged to make things work out best for people on the whole, taking into account human weakness and selfishness. Good rules will channel predominant human motives so that action on them tends to do good. (pp. 17-19)

2. Rules that direct us to take actions found by human experience to be extremely productive of utility are corollaries of the principle of utility. Among these secondary rules are ordinary moral rules such as "Tell the truth," "Keep your promises," and so on. As social life depends on people generally conforming to these rules, we should be taught to conform to them. (pp. 23-24)

3. But rules conflict. Extraordinary situations arise in which acting on a generally useful rule would have bad consequences in this case. We ought to train people not to be slavish "rule-worshippers" but to understand the utilitarian rationale of the rules they obey and to be alert for situations in which maximizing utility requires breaking a rule, even a rule deemed sacred. (See chapter 5)

4. "In the case of abstinences indeed--of things which people forbear to do from moral considerations, though the consequences in the particular case might be beneficial--it would be unworthy of an intelligent agent not to be consciously aware that the action is of a class which, if practised generally, would be generally injurious, and that this is the ground of the obligation to abstain from it" (p. 19, Utilitarianism). This seems to me to diverge from act-utilitarianism toward assertion of rule-utilitarianism, a different doctrine (on which, see below).

5. Along the same line, Mill says at the very end of chapter 2 that people should conform their conduct to secondary rules unless the rules conflict in the particular case. Only in these cases of conflict between rules should people appeal directly to a utility calculation to decide what to do.

6. In chapter 5, on p. 47, Mill writes, "We do not call anything wrong unless we mean to imply that a person ought to be punished in some way or other for doing it--if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience." Mill here identifies the idea of a wrong act with the idea of an act that is fit for punishment. This passage suggests the further view that according to Mill an act is morally obligatory if and only if it would promote utility to punish the failure to do it, and morally wrong only if punishing it in some way would promote utility. This position is a rival to act-utilitarianism. According to Mill, the punishment could consist either in legally enforced penalties, informal social sanctions, or guilt feelings administered by the conscience of the person who does the wrong act. On this view acts can fail to be maximally utility-promoting but not wrong, because punishing them would do no good. When I make a bad selection from a menu at a restaurant, I fail to maximize utility, but this is not wrong according to the p. 47 test. Or suppose people fail to give to charity anywhere near the amount that would maximize utility, but punishing them for uncharity would do more harm than good. In that case the uncharity is not wrong. The idea that Mill is suggesting is that the stringency of moral rules should be adjusted to the weaknesses of human nature. If we try to hold the requirements of moral duty too high, people will balk and duty will fall into disrespect. I think the notion of utilitarianism developed in this paragraph is probably Mill's considered view.

### **Act-utilitarianism and Rule-utilitarianism.**

Act-utilitarianism holds that we ought always to do the act that maximizes utility. Some critics have worried that act-utilitarianism permits too much individual deviation from social rules. Consider the decision whether or not to vote in democratic elections of public officials. An individual vote makes a difference to the outcome only if all the rest of the votes happen to be exactly split, so the individual voter is the tie-breaker. In elections with many voters the probability that one's vote will be decisive in this way is extremely small. Suppose there is a small displeasure associated with taking the time to vote. It begins to look as though act-utilitarian calculation will often yield the decision that one ought to stay home and watch TV rather than vote. Some may consider this understates the obligation of citizens to participate by voting. Or consider wartime rationing of scarce water. If the water in the reservoir stays above a critical level, all will be well. If it falls below that level, the city's water supply will be poisoned. To avert this danger, rules to conserve water are promulgated. Suppose almost everybody is obeying the rationing rules. The act-utilitarian may reason that the water level is not near the critical level, and it is in any case unlikely that one's own water usage will make the crucial difference between clean and poisoned water. So why not violate the rationing rule? A third commonly discussed example is the claim that under certain circumstances act-utilitarianism would appear to endorse the punishment of the innocent. Suppose exemplary harsh punishment of a terrorist offender will deter expected terrorist atrocities, but unfortunately no guilty terrorist is in the government's hands. But an act-utilitarian minister of justice may reason that if we frame an innocent person successfully, potential terrorists will be impressed by our resolve, and will refrain from terrorist activity. So should we send an innocent to the gas chamber? Rule-utilitarianism has been proposed as a version of utilitarianism that preserves its advantages while avoiding endorsement of rule violations in examples like the above. Rule-utilitarians say we

should distinguish two levels of moral thought: the justification of practices and the justification of particular acts falling under practices. At the level of particular acts, we should follow the rules of good practices. We appeal to the goal of maximizing utility in deciding what to do only if we are constructing or revising or justifying a practice.

**Ideal rule-utilitarianism** is the view that we ought always to act in conformity to that set of rules general conformity to which would maximize utility. (Alternate phrasing: We ought always to act in conformity to that set of rules such that everybody's conforming to these rules would maximize utility.) The rationale of rule-utilitarianism is supposed to be that it, unlike act-utilitarianism, clearly supports the obligations we feel to obey useful rules like those mentioned in the examples of the previous paragraph. This doctrine has attracted the objection that it amounts to rule worship. A connected worry is that any version of rule utilitarianism that bids us pay no attention to the actual behavior of others in the situations in which we act but instead asks us to think of a code of rules that would best promote utility if everyone were to conform their behavior to it and then act according to that code involves a bad utopianism. If no version of rule utilitarianism can be devised that is not plagued by these problems, we are back to act utilitarianism and the objections it attracts.

### **Utilitarianism and Consequentialism.**

Utilitarianism is one member of a broader family of views that have been named "Consequentialism." Broadly speaking, consequentialism holds that morality should guide conduct in such a way that the outcome is best. Act-consequentialism is the view that we ought always to act so as to maximize good consequences. In other words, one ought always to choose an act, among the available alternatives, that would produce an outcome no worse than the outcome of any other act one might choose instead. Here doing nothing counts as one among the alternative acts one might choose. This view allows any way of evaluating consequences. For example, we might identify good consequences with nonviolation of a set of individual rights. This "rights" version of consequentialism then directs us to minimize the violation of individual rights. Another version of consequentialism might hold that good consequences consist not just in maximizing utility but also in distributing utility in a fair way. We might, for instance, regard it as a better state of affairs when people get what they deserve according to our preferred notions of desert. Contrary to what Mill argues in chapter 5 of *Utilitarianism*, we might think that justice is a standard of good consequences that is independent of utility. The views just mentioned are (like utilitarianism) members of the family of consequentialist views. The question arises whether some or any of the difficulties and objections that plague utilitarianism also attach to the more general and abstract idea of consequentialism. In this course, Amartya Sen defends a consequentialist but nonutilitarian position. Here I simply list some issues that need to be decided in the course of assessing utilitarianism.

1. Utilitarianism versus fairness. Contrary to common notions of fairness, utilitarianism seems to permit us to be "free riders" or parasites on useful schemes of social cooperation when others are doing their share. Also contrary to common notions of fairness, utilitarianism requires us to contribute to schemes of cooperation even when others are not doing their share, if one's individual contribution does more good than any alternative act open to the agent.

2. **Agent-relativity.** Some moral views hold that what an agent is morally required or permitted to do may vary depending on the agent's relation to those whom his acts might affect. On such a view, I may have a special obligation to my wife and children, so that if the ship is sinking and lifeboats are scarce, I am right to grab the last boat for my family, even though I could have saved more people by putting another family in the last boat. An agent-relative moral theory gives different agents different aims. An agent-neutral theory gives all agents the same aims. Utilitarianism is an agent-neutral theory. It gives all agents the one aim of doing whatever maximizes utility. Special obligations to one's family, one's friends, one's nation, one's gender, one's species, and so on are problematic in a utilitarian perspective. (Here's an alternative characterization of agent-relativity. If any proper statement of a reason makes an essential reference to the agent who has the reason, the reason is agent-relative. Example: "You should keep the promises that you make" states an agent-relative reason. "Each person should always do whatever would maximize the fulfillment of her own interests" states an agent-relative reason. The underlined personal pronoun in the formulation of the reason is ineliminable. In contrast, a reason that can be properly stated without making essential reference to the agent who has the reason is an agent-neutral reason.)

3. Absolutism regarding moral rules. No consequentialist view can endorse an absolute, exceptionless obligation to obey any moral rule (except "maximize good consequences"). Suppose it is agreed that the right not to be murdered is the most sacred right there is. Nonetheless, we may find ourselves in a tricky situation such that two persons will certainly be murdered unless we ourselves murder one. If good consequences are held to consist above all in avoiding murders, then consequentialism will say we ought to commit one murder to avoid two.

Involved here is consequentialism's commitment to what has been called the Negative Responsibility thesis: that insofar as we are ever morally responsible for anything, we are just as much responsible for events that we allow to happen or fail to prevent as for events that we ourselves bring about by our actions.