

Introduction to Rawls on Justice and Rawls on utilitarianism.
For THEORIES OF JUSTICE USD Fall, 2008
Richard Arneson

In chapter 1 of *A Theory of Justice* John Rawls introduces the conception of justice that he affirms—'justice as fairness,' a version of social contract theory in the tradition of Locke, Rousseau, and Kant. He sees two main rivals to his approach—intuitionism and utilitarianism.

According to Rawls, the principles of justice are principles that determine a fair resolution of conflicts of interest among persons in a society. "A set of principles is required for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares" (p. 4). Different interpretations or conceptions of justice fill out this core concept; a theory of justice seeks a best conception. Justice takes priority over other normative claims—as Rawls states, justice is the "first virtue of social institutions." The principles of justice are principles to regulate what Rawls calls the "basic structure of society," that is, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation" (p. 6). Why focus on the basic structure? "The basic structure is the primary subject of justice because its effects are so profound and present from the start. The intuitive idea here is that this structure contains various social positions and that men born into different positions have different expectations of life determined, in part, by the political system as well as by economic and social circumstances. In this way the institutions of society favor certain starting places over others. These are especially deep inequalities. Not only are they pervasive, but they affect men's initial chances in life; yet they cannot possibly be justified by an appeal to the notions of merit or desert. It is these inequalities, presumably inevitable in the basic structure of any society, to which the principles of social justice must in the first instance apply" (p. 7).

Justice as fairness is the idea that the principles of justice are the ones that "free and free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association" (p. 19).

Intuitionism is the "no-theory theory." It says there are several principles of right and none invariably takes priority over the others. No theory of justice is correct, a theory being a set of principles that (1) singles out the empirical facts relevant for choice and (2) when these relevant facts are fully specified for any decision problem, the principles then determines what choices should be made, what actions or policies ought to be selected (what institutions and practices established).

Utilitarianism is the doctrine Rawls regards as the most formidable version of teleology, which he defines as a theory that (1) defines the good as independent from the right and (2) identifies the right with maximizing the good. Taken as a theory of justice, utilitarianism holds that "society is rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it" (p. 20). Rawls objects that utilitarianism ignores the separateness and distinctness of persons and does not recognize that justice is what free persons would choose as the principles to regulate their social cooperation under conditions that are fair. Question: What exactly is the separateness of persons objection against utilitarianism?

The virtue of prudence requires an individual to be impartial in making decisions that affect all of the times of her life. A benefit at one time should count the same as an identical benefit that could instead be gained at another time, and the same with losses. Rawls notes that utilitarianism extends the required impartiality across all persons. A benefit or loss for a person should count the same, in the determination of what should be done, as a same-sized benefit or loss that would befall any other person at any other time or at any other place. The extreme impartiality built into utilitarian thinking may sound noble or ideal, but Rawls finds it problematic.

Rawls treats utilitarianism as a candidate theory of justice for the regulation of major institutions. It is usually taken to be first and foremost a theory of right action. Broadly speaking, utilitarianism holds that morality should guide conduct in such a way that the outcome is best for people on the whole. This might be interpreted as

Act-utilitarianism = one ought always to do that act which, compared to available alternatives, maximizes utility. Act-utilitarianism so understood is a test or criterion of what one should do (a test of right action, one may say). It is not necessarily a guide to decision making, and indeed cannot directly serve as a guide to decision making when one does not know which of the acts that one could perform will maximize utility. Associated with act-utilitarianism is an ancillary test (a guide for decision making) intended for use when one knows the value of each of the outcomes that could result from one's actions and can estimate the probability of any given outcome occurring if one does an act that might lead to it. Expected-utility maximization = one ought always to do the act which, compared to available alternatives, maximizes expected utility. Here the expected utility of an action is the sum of the value of its disjoint outcomes times the probability of each one's occurrence.

The rationale of expected utility maximization is that if a decision problem were repeated many times, expected utility maximization would produce more utility than would following any other decision rule. (It is not uncontroversial, however.)

In general, the act utilitarian needs to distinguish the objectively right act, the act that of the given alternatives would actually maximize utility, and the subjectively right act, the act that on the basis of the information available to the agent at the time of choice is the one that is most rational to choose for one whose goal is to maximize utility.

We can also distinguish **average utilitarianism and total utilitarianism**. The former says one ought always to act so as to maximize average utility (utility per person). The latter says one ought always to act so as to maximize total or aggregate utility. The two views come to the same unless one is making decisions that will affect the number of persons in the world, such as deciding whether or not to have a baby or deciding among social policies that will encourage or discourage a higher birth rate.

Utilitarianism versus Common-sense Morality.

Utilitarianism appears to conflict with what we might call "common-sense morality," the view that takes morality to be constituted not by any goal to be pursued but by rules to be followed. The rules of common-sense morality by 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.

Utilitarianism conflicts with common-sense morality in denying options as well as in denying constraints. If one has moral options, there are actions available for choice, all of which are innocent and eligible for choice, and none of which it would be morally wrong to choose. We typically think that so long as one does not violate certain moral constraints (don't lie, don't steal, don't kill), one under most circumstances is morally free to choose among many different course of action and ways to live, whereas utilitarianism rejects this picture.

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.

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. 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 (*Anarchy, State, and Utopia*, chapter 3).

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.

Hybrid views are possible. One might hold that the good life consists of the enjoyment of the excellent (combining hedonism and perfectionism). 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 pleasurable. 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.

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. (Following suggestions by Richard Brandt, Brad Hooker has devised a version of the view that aims to be responsive to these difficulties.)

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.

Rawls & utilitarianism.

Utilitarian recommendations for conduct are highly sensitive to circumstances. For a hunter-gatherer culture in harsh circumstances such as the Inuit in the far north circa 1900, utilitarianism might endorse such practices as enforced euthanasia of aged persons and the required infanticide of female children born to a household that lacks a male child, whereas for Californians in mild circumstances circa 2008, these practices would not be endorsed.

It will turn out that Rawls proposes both a general conception of justice that is valid anywhere and everywhere and a special conception for modern times. Rawls also considers a simplified version of the problem of selecting principles of social regulation. He considers the case of a single society isolated from others and asks what is just for such a setting. He also limits himself to developing a theory of justice for a "well-ordered society," one in which "(1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles" (p. 4). So problems that call for a criminal justice system are set aside, along with problems of deciding how to act in an unjust society and how to improve it. These simplifications are made mainly to make the difficult problem of figuring out what is just more tractable, but publicity is also a moral requirement for Rawls: it would be morally wrong not to make accessible to all members of society the basic norms by which it is run and their rationale. For the utilitarian, publicity might be good or bad depending on the utility consequences of implementing it.

Finally, as Rawls notes, justice as fairness and utilitarianism are not, abstractly considered, strictly inconsistent. The principles free and equal rational persons would choose under fair conditions might for all we have said so far turn out to amount to utilitarianism. Rawls will argue that in fact this is not so.