

JOHN RAWLS'S THEORY OF JUSTICE
NOTES FOR PHILOSOPHY 167 Dick Arneson SPRING, 2005

CHAPTER ONE.

Rawls and the Social Contract Tradition

Rawls aims to develop a theory of justice that will be superior to utilitarianism and that will supplant what he calls "intuitionism" (the No Theory theory). According to Rawls, a *moral theory* is a set of principles that (1) stipulates what information we need in order to decide what to do and (2) determines what should be done in any circumstances, provided we have the information regarding those circumstances that the principles themselves specify to be relevant. In other words, no further evaluation is required; the principles embody the evaluation needed to identify morally right policies. With a theory, given a specification of the relevant facts and a statement of the principles, one can derive as conclusion what should be done.

But if one's morality includes more than one value, how can one avoid the need to weigh these plural values against each other intuitively on a case by case basis? This is what Rawls calls the "priority problem." To solve it we need to build in a weighting or priority ranking of the different values we accept as relevant into the formulation of our principles. Rawls favors what he calls "lexical priority rankings." If one value has lexical priority over another, the first one trumps the second, we should do everything we can to achieve the top-ranked value to the greatest degree possible and devote resources to achieving the lower-ranked value only when doing so does not lessen even in the slightest degree the extent to which we achieve the top-ranked value.

Rawls proposes to develop a theory of justice by revising the social contract tradition of theorizing about justice associated with the 17th and 18th century writers John Locke, Jean-Jacques Rousseau, and Immanuel Kant. Locke sees legitimate political authority as deriving from the free and voluntary consent of the governed, from a contract or agreement between governor and governed person. Rawls says he will take the social contract idea to a higher level of abstraction. According to Rawls, justice is what free and equal persons would agree to as basic terms of social cooperation in conditions that are fair for this purpose. This idea he calls "justice as fairness." The conditions that Rawls takes to be most appropriate for the choice of principles of justice constitute what he calls the "original position."

Rawls construes the task to be choosing principles for a "well-ordered society," a society that is (a) effectively regulated by a public conception of justice and (b) whose members understand and give allegiance to this public conception. Moreover, a third condition, (c), holds: it is common knowledge among all members of society that a and b hold. Why this idealization? Rawls thinks we

need to get clear about first-best theory before we can be in a position to think through problems that arise when institutions are not just and some persons are not disposed to comply with requirements of justice.

The Basic Structure of Society as the Primary Subject of Justice.

Rawls writes, "For us the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation. By major institutions I understand the political constitution and the principal economic and social arrangements. Thus the legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family are examples of major social institutions" (p. 6). Why focus on the basic structure? Rawls: "Because its effects are so profound and present from the start." He says that the basic structure brings it about that people are born into different social positions with different and unequal life prospects. These initial inequalities are "especially deep," says Rawls. They make a big difference to our lives but lie beyond our power to control—a baby cannot choose her parents or their social position. Rawls is implicitly contrasting deep inequalities of this sort with shallow inequalities, the kind that arise via the choices and actions of individuals, for which they might be deemed responsible. Rawls concludes: "It is these [deep] 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).

{Criticism: Deep inequalities as characterized by Rawls do not all have their origin in the basic structure of society. The differences in native talent or talent potential between people are also present at birth and fundamentally shape people's life prospects, but these arise from different genetic endowments not from the basic structure of society.}

Rawls versus utilitarianism.

Rawls writes, "Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override" (p. 3). He also writes, somewhat cryptically, "Utilitarianism does not take seriously the distinction between persons" (p. 24). What is he saying?

Utilitarianism says, one morally ought always to choose the act or policy that maximizes utility. Utility is a measure of individual good. Rawls considers the measure to be informed desire satisfaction. So if one faces a choice between actions, for each action, determine the impact that action would have on the utility of every person who would be affected by it, sum the results, and pick the action that would lead to the highest utility total (the lowest negative sum if all choices lead to negative utility). If the actions one might take will affect the number of people who shall live, then two versions of utilitarianism may disagree. Average utilitarianism says choose the act that maximizes utility per

person; aggregate utilitarianism says choose the act that maximizes the sum of utility across persons.

Given constant population, utilitarianism says we should pick the act that leads to the highest sum of utility regardless of how this utility is distributed across persons. So in the example, utilitarianism is indifferent between policy 1 and policy 2. From the utilitarian standpoint, one is as good as the other.

	Policy 1	Policy 2
Utility outcomes for the individual given choice of the policy		

Sam	20	7
Sally	-10	6
Samantha	10	7

The sum is +10 in each case, so according to utilitarianism, it is a matter of moral indifference, which policy we choose. But some might say we care not merely about the total of utility but about how it is distributed across persons. Another example:

Utility outcomes for Individuals given choice of the Policy	Policy 1	Policy 2
Rich	14.999	8
Poor	00.001	7

(Assume equal numbers of Rich and Poor)

Again, since the sum of utility produced by 1 and 2 is the same, utilitarianism is indifferent between these policies.

Rawls thinks the mere fact that losses imposed on one individual are in utilitarian calculation overbalanced by aggregate gains to other people does not in and of itself suffice to show that it is morally acceptable to push one to the wall for the sake of gains to others. Each person has her own life to live, one might say.

Or consider Jim Crow laws. The utilitarian will oppose such laws just in case the utility gains that accrue under these laws to the favored whites are overbalanced by losses to blacks who suffer discrimination. If there is a legal policy other than Jim Crow that in the circumstances would produce greater net gains, utilitarianism favors the alternative and rejects Jim Crow. But the utilitarian right answer depends on iffy calculation of consequences. Rawls says, even if the utilitarian comes out with the right answer, choice of acceptable social policy, in such decision problems, the utilitarian would be getting the right answer for the

wrong reason. Intuitively the blacks in this situation have equal rights, and the gains the whites could get by violating these rights should not get counted as a “plus” in favor of policies of discrimination.

Utilitarianism is an example of a teleological theory, according to Rawls. A teleological theory is one that (a) defines the good independently of the right and (2) identifies the right with maximization of the good. In justice as fairness, by contrast, the principles of right (morality, justice) are determined to be what they are pretty much independently of what is good. Once the principles of right are at hand, and regulate society, each individual should develop a conception of the good and a plan of life that does not require violation of rights for its fulfillment. A person’s satisfactions taken in violation of other people’s rights should not count positively at all in the determination of just social policy.

Some remarks on moral theory; reflective equilibrium.

How would we go about deciding which of competing moral conceptions is better, worthy of acceptance? Rawls suggests that at any given time we have both particular beliefs about what should be done in particular circumstances and general beliefs for or against various principles. At any given time the set of our moral beliefs is probably inconsistent—some of our beliefs contradict other beliefs we also hold. In moral argument we seek a reflective equilibrium—a position in which our considered particular judgments that we want to affirm can be derived from general principles we find intuitively attractive, the entire set being internally consistent. Our beliefs are then in reflective equilibrium, equilibrium after reflective scrutiny. This equilibrium can be disturbed by further moral thoughts and arguments. (If I hold that abortion is OK but infanticide is wrong, but then cannot answer the challenge to show how a newborn baby is different in any morally relevant respect from a nine-months-old fetus in the womb, I am thrown into reflective disequilibrium.) We can then stipulate that wide reflective equilibrium is what we would believe after subjecting our considered particular and general judgments to all the arguments we can discover at a time, and ideally extended reflective equilibrium is what we would believe about morality at the limit of moral inquiry, after full reasonable scrutiny considering all pertinent arguments. Rawls hopes to take a few steps toward reflective equilibrium but does not anticipate reaching it.

CHAPTER TWO. TWO PRINCIPLES OF JUSTICE

Rawls has two big ideas. One is that the principles of justice are whatever would be chosen by free and equal persons as basic terms of social cooperation under conditions that are fair. Chapter 3, on the original position, works out this idea.

Rawls’s second big idea is a vision of a just society, encapsulated in two principles. He thinks these two principles are the basic terms of social cooperation that would be chosen in the original position. That they would be

chosen in the original position counts as a justification of them. Another justification of the principles is that, when we think about it, their implications for policy match our considered reflective judgments on these matters. Rawls introduces and characterizes these principles in chapter two.

Primary social goods.

Rawls associates the idea of liberalism with the idea that conceptions of what is good and worthy of pursuit in human life cannot provide a shared basis for social cooperation in modern society. In this perspective, liberalism grows from and generalizes the norm of religious toleration as it developed during the wars of religion in Europe after the Protestant Reformation. Religion cannot be the basis of social unity, because rational people will disagree on matters of religious doctrine and over time there is no discernible tendency toward any consensus. And just as we worship different gods, we seek different goods. Some people seek business and entrepreneurial success, some people seek a state of harmonious community with the natural environment. We aren't ever going to agree about such issues. We have to agree to disagree. Good is subjective: what makes something fundamentally good for you is that you desire it, and equally rational people have different basic desires.

It may seem that Rawls has painted himself into a corner. If we can't agree on what is good, how can we agree on principles of justice that regulate the distribution of fruits of social cooperation—the goods generated by social cooperation? Rawls proposes is that even though we disagree about what is ultimately good, we may agree that there are certain general purpose means that will be useful to carrying out a wide variety of different plans of life. Consider the standpoint of a young adult. She has some conception of her good, what she thinks worthy of pursuit in life. But she knows her values and aims might well shift as she goes through life. Seeing this, and wanting to give herself a reasonable chance of satisfying her life aims, whatever these might turn out to be, she has reason to want certain general purpose goods—things that any rational person will want, whatever else she wants. These are the primary goods. Some of these primary goods are social primary goods—they are distributable by social arrangements. Primary goods provide a basis for interpersonal comparisons that is suitable for the theory of justice. From the standpoint of social justice, we say people are better off or worse off, depending on their holdings of primary social goods. Elsewhere in his writings Rawls defines primary social goods differently: as those things any rational person who gave priority to developing and exercising her capacities for a sense of justice (cooperating with others on terms agreed to be fair) and for a conception of her good would want, whatever else she might want.

Rawls writes, "The primary social goods, to give them in broad categories, are rights, liberties, and opportunities, and income and wealth" (p. 79). He

immediately adds, "A very important primary good is a sense of one's own worth, but for simplicity I leave this aside until much later, section 67." [In 67 Rawls states that self-respect is perhaps the most important primary good, and it has two aspects: "it includes a person's sense of his own value, his secure conviction that his plan of life, his conception of his good, is worth carrying out. And second, self-respect implies a confidence in one's ability, so far as it lies within one's power, to fulfill one's intentions."]

Elsewhere in Rawls's writings he provides this specification of the primary social goods. The list includes: "a. basic rights and liberties, also given by a list, b. freedom of movement and free choice of occupation against a background of diverse opportunities, c. powers and prerogatives of office and positions of responsibility in the political and economic institutions of the basic structure, d. income and wealth, and finally, e. the social bases of self-respect."

So we are to decide whether people are being fairly treated by society by examining their access to primary social goods. Why not instead look directly at the quality of life people gain, by whatever they do with their primary social goods? Rawls affirms at this point a certain norm of personal responsibility. Given that the distribution of primary social goods is fair, it is up to each person, each person's responsibility, to try to fashion a satisfying life for herself. How well your life actually goes in the just society is your business, not the proper business of society. Against the idea that society should rate individual lives according to the satisfactions individuals get, Rawls writes, "Justice as fairness, however, takes a different view. For it does not look behind the use which persons make of the rights and opportunities available to them in order to measure, much less to maximize, the satisfactions they receive. Nor does it try to evaluate the relative merits of different conceptions of the good. Instead, it is assumed that the members of society are rational persons able to adjust their conceptions of the good to their situation" (pp. 80-81). Just as under religious toleration your salvation is your own business, not the job of the government, under Rawlsian justice your quality of life (given a fair distribution of primary social goods) is your own business, not the responsibility of society or government.

Two principles of justice.

In chapter two Rawls gradually introduces his principles by proposing more general formulations and then trying to show that these formulations can be variously interpreted at different points, and that the result of inserting the best interpretations of the initial vague formulations at each point of decision gives you his favored principles. At the end of this process we get these principles:

1. Equal Liberty. "Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all" (p. 220).
2. Social Inequality. "Social and economic inequalities are to be arranged so that they are both (a) to the greatest expected benefit of the least advantaged and (b) attached to positions and offices open to all under conditions of fair equality of opportunity" (p. 72).

The "basic liberties" in the first principle are specified by a list: "political liberty (the right to vote and to hold public office) and freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person, which includes freedom from psychological oppression and physical assault and dismemberment (integrity of the person); the right to hold personal property and freedom from arbitrary arrest and seizure as defined by the concept of the rules of law" (p. 53). These might be thought of roughly as constitutional liberties—the sorts of fundamental liberties that might be protected by the political constitution of a modern democratic society. The equal basic liberties specify a status of equal democratic citizen that is to be accorded to all members of society.

Notice that the basic liberties do not include every liberty that might be valued. One might wonder, what about the freedom to do what I please so long as I don't harm others? How wide this freedom of action should be, according to Rawls, is left to the discretion of a democratic legislator. Different legislatures in different equally just societies might trade off nonbasic liberties against other benefits in various ways, without compromising the justice of the basic structure.

Economic liberty has a somewhat uncertain status in relation to Rawls's list of basic liberties. At one extreme, a command economy in which the state assigns people to economic roles and jobs is ruled out as inconsistent with basic liberty (see section 42 in chapter 5). Presumably the free choice of occupation (anyone may apply for any job) and the decision whether to apply for part-time or full-time work and whether to take larger or smaller leaves of absence from paid employment or paying self-employment over the course of one's life are matters of basic liberty, though not explicitly mentioned by Rawls. One has the right to own personal property, but the extent to which major means of production should be in private hands or publicly owned is not a matter of basic liberty according to Rawls.

The first principle of justice, Equal Liberty, has lexical priority over the second principle. This means one must not accept any tradeoff of lesser fulfillment of the first principle for greater fulfillment of the second—no trading lesser freedom of speech for greater economic gains, for example. [A somewhat more

complicated statement of the lexical priority rankings among Rawls's principles occurs on pp. 266-267.]

Within the second principle, there is another lexical priority specified: Part b, fair equality of opportunity (FEO), has lexical principle over part a, otherwise known as the difference principle.

The difference principle (part a of the second principle).

The difference principle says that social and economic inequalities are to be arranged so that they are to the greatest expected benefit of the least advantaged. Expected benefits are measured by expected individual holdings of primary social goods (other than the basic liberties, which are regulated by the first principle).

First question: Why doesn't the idea of making inequalities maximally advantageous for the people who have least straightforwardly imply eliminating all social and economic inequalities? After all, in dividing any fixed stock of goods, the distribution that makes the least share as large as possible is equal division.

The key here is that the stock of goods is not fixed. So to speak, how we divide up the pie at one time affects people's incentives to make economic contributions and hence the size of the pie at later times. If the rule is that economic goods will always be divided equally among all persons, economic production in a large society may well shrink—people find ways to avoid contributing, since whether they contribute a lot or a little or not at all will not affect the share they get. The right rule, according to Rawls, is the rule or set of economic arrangements that maximizes not the relative share but the amount of primary goods that goes to the worst off group over the long haul. (A complication here appears in chapter 5, where Rawls considers that there is a separate issue of intergenerational savings, which interacts with the implementation of the difference principle. See chapter 5.)

For simplicity, and just to get the flavor of Rawls's approach, suppose we are dealing with a private market economy in which there is taxation of earnings, the funds to be transferred in ways that increase the overall primary goods holdings of the worst off. The difference principle in this setting would require that the tax rate be set so that the long-run benefits to the worst off are maximized. (Conceivably this tax rate could be anywhere between zero per cent and 100 per cent.) In other words, some inequalities in social and economic benefits are productive in the sense that arranging the economy so that these inequalities are generated works to maximize the long run primary goods benefits flowing to the least advantaged.

Why does justice require **maximizing** the benefits that go to the very worst off? Why does justice require tilting in that extreme way in favor of the people who happen to end up at the bottom? Rawls's full answer is supposed to come in chapter 3, the original position account of choice of principles of justice.

Another complementary answer is suggested in chapter 2. This answer is further developed in chapter 5, sections 47 and 48. Suppose we have a market economy with private ownership, and someone proposes that the outcome of free market trading is fair. "I deserve and merit the extra advantages that come to me through my hard work, entrepreneurial or risk-taking or whatever. Why does justice require taking from the ants (the hard workers in the fable) and giving what they produce to the grasshoppers (who just play around all summer and then have no food when winter arrives)?" Suppose someone makes this argument. Rawls's answer is that what we might deserve and reasonably take credit for in our own efforts is inextricably tangled up with factors that are beyond our power to control and that we cannot reasonably take credit (or discredit) for. We don't deserve our genetic endowment that gives us the potential traits that will make us successful or unsuccessful in market competition. Nor do we deserve our favorable or unfavorable early socialization. These are matters of sheer luck that just befall us, but they deeply affect our prospects for market success. It's not that no one ever deserves anything for her ingenuity and efforts, but that what is sheer luck and what is really due to us are mixed together and inseparable so that, Rawls, writes, "The idea of rewarding desert is impracticable" (p. 274). Moreover, a market economy is misunderstood if it is regarded as a set of institutions designed to reward the deserving. The market economy is responsive to supply and demand conditions, and when it is working well the desirable trait it manifests is Pareto efficiency (a state of affairs is Pareto efficient when one cannot alter the state of affairs to make someone better off without making someone else worse off), not distribution of reward according to true deservingness. Since the extra advantages that the better off gain in market competition are not deserved in any deep way, there is no bar to taking some of them to improve the condition of those who are worse off. The logical limit of this procedure is the difference principle.

Another way to put Rawls's reasons for favoring the difference principle. We start with a presumption in favor of equal distribution—since no one is more deserving really than anyone else so far as we can know, we suppose everyone should get the same level of what goods and benefits there are to be had. We relax the presumption in favor of equality when inequalities would make us better off. If our initial preference for equality reflected special solicitude for those who would end up worst off under inequality, then the natural way to pick out which system of inequality is morally best, is to single out the system that is maximally advantageous for those who end up worst off.

Fair equality of opportunity.

We can distinguish two ideas of equality of opportunity. One is formal equality of opportunity or careers open to talents. It is satisfied when competitive positions of advantage (good jobs, opportunities to borrow venture capital, places in universities and colleges) are open to all applicants and applications are judged on their merits. The merit of an application is the degree to which its selection would be reasonably expected to further the morally innocent purposes of the enterprise. So under formal equality of opportunity, banks lending venture capital are open to all applicants, and they pick the applications for money that are best in the sense of promising the greatest return on its loan to the bank. Under formal equality of opportunity, a business firm that wants to hire someone for a position accepts applications from anyone for the post, and the best applicant is selected—the one whose hiring will do most to advance the morally innocent purposes of the enterprise (often, to make profits).

Notice that a society could satisfy formal equality of opportunity even though only people born to wealth and high status gain access to the education and socialization that renders them qualified for competitive positions, so only that segment of the population ever becomes successful. If we institute public schools financed by general taxation, we are in favor, to some extent, of providing member of society some opportunity to become qualified for competitions for especially lucrative or desirable positions. A state-financed program that provides pre-school training to disadvantaged young children is another step in the same direction.

A substantive principle of equality of opportunity demands more than formal equality of opportunity, and what it demands is fair distribution of opportunities to become qualified for social competition. Rawls's proposed principle of fair equality of opportunity (FEO) expresses the extension of this norm to its logical limit. When FEO is satisfied in a society, any two people with the same native endowment of talent and the same ambition (the same willingness to work for success) will have identical prospects for competitive success. In other words, when FEO is satisfied, being born in one or another social position, to rich or poor parents, well educated or poorly educated parents, being born with one or another skin color or ethnic background—none of these factors by themselves influences anyone's prospects for competitive success (except insofar as any of them might influence the factors that FEO singles out as legitimate causal determinants of prospects for success). Being born on the right side or the wrong side of the tracks does not matter to my life prospects, in the society effectively regulated by FEO. If I have the same native talent and same ambition as someone born to social privilege, then the fact that I am born in disadvantaged circumstances will end up making no difference to my life

prospects. A Rawlsian society that satisfies FEO might be viewed as a perfect meritocracy.

Rawls says that social inequalities are morally acceptable just in case two requirements are satisfied. One is that the positions that give access to the long end of the stick, the better than average payoff, are open to all under conditions of FEO. The second condition is that these inequalities must be set so that they work to produce the maximal benefits that accrue to the worst off over the long run. The FEO principle has lexical priority over the difference principle.

You might think of Rawls's principles as expressing ideals that he sees as part of the public culture of modern constitutional democracies. The Equal Liberty Principle reflects the idea that fundamental rights such as rights to free speech and democratic citizenship are to be guaranteed to all even against majority will. Fair Equality of Opportunity reflects the ideal of a society in which social class and racial and caste hierarchy are entirely overcome. The difference principle carries to the limit the idea that perhaps finds public expression in the New Deal in the U.S. in the 1930s, the idea that the market economy should be regulated to bring about a fair deal for those who lose out in free market competition through no real fault or choice of their own but rather through random unchosen luck.