

Rawls' Theory of Justice

The goal: explain how justice requires us to order the *basic structure* of society—its major institutions and social arrangements.

Justice as fairness: Rawls' fundamental claim is that justice consists in those principles that we would choose collectively from a position of fairness. This is in line with the social contract tradition (Hobbes, Locke, Rousseau, Kant).

The critique of utilitarianism (U): U as a theory of justice tells us to arrange institutions so as to bring about the greatest good for the greatest number. That is, it defines the Right (what we or the state should do) in terms of the Good (desirable outcomes). Rawls, by contrast, insists that the Right has priority over the Good: which outcomes we can produce is constrained by the need to adopt appropriate means.

Key argument: U makes sense as an *intrapersonal* theory: sometimes it makes sense to trade off some parts of your life for the sake of making the whole of your life better. But when U is applied as an *interpersonal* theory it becomes implausible. You cannot trade off the good of some for the sake of making others better off. To do so is to *ignore the distinction between persons*.

The Two Principles

- (1) Each person is to have an equal right to the most extensive basic liberties compatible with a similar liberty for all.
- (2) Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged (the Difference Principle), and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.

These principles are to be *lexically ordered*: the first must be satisfied before we can try to satisfy the second; trade-offs are forbidden.

The Two Lotteries: Rawls intuitively motivates the Second Principle by pointing out that the alternative is to leave people at the mercy of the social and the natural lotteries—how we fare will be a matter of luck. Rawls suggests justice requires us to neutralize that sort of luck. Leaving people to the luck of the draw will produce morally arbitrary outcomes that are unfair or undeserved. Rawls might thus be seen as a *luck egalitarian* (though this is controversial).

The Original Position: the position, analogous to the classic state of nature, from which Rawls asks us to imagine ourselves deliberating about which principles of justice to choose. Since he thinks all rational people will converge on the same principles, we can imagine everyone or just a single person chosen at random doing the deliberating. Principles are to be chosen on the basis of rational self-interest. The goal is to maximize access to *primary goods*—the goods we always want more of to realize our goals.

The Veil of Ignorance: deliberators in the OP know basic facts about human psychology, economics and the like, but are completely ignorant of their particular interests, goals, preferences, social status, natural abilities, etc. The theory enforces impartiality, since our ignorance prevents us from taking advantage of special features of our own situation. (Cp. the categorical imperative.)

Maximin reasoning: Rawls says we'd pursue something like maximin reasoning in the OP; that is, we'd try to maximize the minimum outcome. This is a conservative policy that we'd be inclined toward if our enemy were going to choose our social position. It is motivated by the fact that we will be required to live with the principles we choose, and so will be unwilling to pick principles that risk leaving us in an intolerable position.

Against Average U: Rawls denies we'd choose average-U in the OP. Average-U tells us to arrange institutions so as to maximize average well-being. Rawls objects that this assumes a Laplacian indifference principle—assumes, that is, that all the possible outcomes in a society are equally likely. For if some outcomes were more like than others, then we couldn't make an average-U calculation. But in the OP we're barred from any such probability estimate. (Example of the money game: you pick one bill from either of two pots, one of which has 100s and 20s, the other of which has 100s and 1s; for all you know, the ratio of 100s to 20s is 1:1, while the ratio of 100s to 1s is 100:1.)

An objection: suppose the possible outcomes are the following columns. (Numbers represent how well off you'll be.) You can choose among strategies D1 and D2.

D1: 0 N
D2: 1/N 1

D2 is the maximin strategy: for any N, the worst-case for D1 is 0, whereas for D2 it's 1/N. Suppose N is low. Then D2 might make sense: the most you can get under D1 is 2 (say), and the difference between 0 and ½ might be significant. But suppose N = a million. Then it seems crazy to ignore the maximum under D1, and the difference between 0 and 1/1,000,000 is trivial anyway.

A related objection: suppose we're now at distribution A1, and choosing between policy options that will produce A2 or A3. The numbers represent how well off two groups are, the Haves and the Have-nots.

A1: 100 10
A2: 90 12
A3: 35 15

If we arrange things so they are best for the worst off, we'll implement A3. But maybe that's a mistake. A2 isn't much worse for the worst off, and massively better for the better off. (Imagine the worse off compose a tiny fraction of the population.)