

THE ORIGINAL POSITION PHILOSOPHY 167 R. ARNESON
A Brutally Short Summary

These pages consist of exposition except for occasional interspersed criticism and commentary. These passages of criticism and commentary are marked by being enclosed in double brackets, like this. [[Rawls is full of beans.]]

John Rawls holds that the principles of justice are what free, rational, mutually disinterested persons would choose (in an initial situation of equality) as basic terms of cooperation to regulate a society they expect to inhabit. The *original position* is his name for the philosophically most favored interpretation of this choice situation.

Rawls advances two types of arguments for his proposed principles of justice: (1) they match our considered judgments about what justice requires in various particular circumstances, and (2) they would be chosen in the original position. The original position is supposed to register our intuitive judgments concerning the circumstances that render the choice of moral principles a fair choice.

[[Two generic objections to Rawls's social contract or original position argument: (1) given his set-up of the choice situation, Rawls's favored principles would not be chosen. (2) Whether or not his principles would be chosen, hypothetical choice by these imaginary persons does not give us reasons to favor his principles. Actual consent obligates; hypothetical consent does not. (Another version of 2.) We do not in fact have views about what conditions constitute fair conditions for choice of moral principles. We have views about fair bargaining conditions, and about fair procedures for applying principles already on hand, but these are different matters. The right conditions for choice of principles are whatever conditions would lead to choice of correct principles. So the original position argument has no independent justifying force; at best it is a roundabout device for presenting reasons that could be stated directly.]]

Sections 20-25, the set-up of the choice situation.

1. The circumstances of justice. Absent these circumstances, the concept of justice either does not apply or has no point. Rawls, p. 110: "the circumstances of justice obtain whenever persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity." Mutually beneficial cooperation is possible and not otiose. Individuals all have their own individual aims; they are not united by a conception of a common good like members of a monastery.

[[It follows that on Rawls's view, if a ship is sinking and there are not enough life preservers for all and one life preserver saves exactly one person, the concept of justice does not apply, because the circumstances of justice do not obtain.]]

2. The constraints of the concept of right. These are constraints that hold for the choice of all moral principles. They are—

Generality—principles are formulated in general terms, without names or rigged definite descriptions. "Give all the money to Arneson" is not a candidate moral principle.

Universality—Principles are to be universal in application. Rawls: "Principles are to be chosen in view of the consequences of everyone's complying with them.

Publicity—the principles chosen are to be part of a public conception of justice. Anyone in society knows or can know the content of the principles, and anyone can also know that the condition in the previous clause of this sentence holds. Ruled out by publicity is an esoteric conception of justice, say one known to a colonial governing elite but hidden from the governed subjects.

Ordering—the conception chosen "must impose an ordering on conflicting claims." For any set of circumstances, the principles chosen specify what justice requires in those circumstances.

Finality—"The parties are to assess the system of principles as the final court of appeal in practical reasoning." Everyone complies with the verdicts reached by this final court of appeal.

3. The veil of ignorance. Those in the original position are choosing principles while under a thick veil of ignorance. They know no particular facts about themselves, only such general facts

as science and social science provide. Particular acts include facts about one's own personal traits, one's ethnicity and race, one's gender, one's religion and moral beliefs, one's personal values, and one's location in history. The veil of ignorance deprives those choosing principles of information that might give anyone a bargaining advantage or enable a person to tailor a principle to her own advantage. (For example, in the original position, you don't know whether you are rich or poor, so would have no motivation to propose principles that favor rich or poor.)

4. The rationality of the parties. Those in the original position seek to advance their interests as best they can by their choice of principles. They are choosing principles to regulate a society they expect to inhabit, and they choose principles based on how they expect to fare in a society run by the principles chosen. Rawls assumes the parties in the original position choose rationally to advance their interests.

Rawls stipulates that the parties are "mutually disinterested." This means each assumes she has a set of aims she wants to fulfill, and she takes no interest in the fulfillment of other people's interests *per se*.

Problem: How can I advance my interests if I am wearing the veil of ignorance and I have no idea at all what my interests are? Rawls: The parties know that whatever their interests are, they will benefit from having primary social goods. Primary goods are ones it is rational for anyone to want, whatever else she wants. Primary *social* goods are primary goods that can be distributed by society. Good health, though primary, is not a primary social good. The parties in the original position choose principles based on the extent to which they can expect to gain primary social goods under the rule of those principles.

The parties are assumed by Rawls to be nonenvious. Each cares only about the amount of goods she expects to get under the rule of proposed principles rather than about her relative share.

On primary goods, see section 15 of chapter 2, and also the preface to the revised edition. In this preface, Rawls stipulates that primary goods are goods it would be rational for anyone to want who has two moral powers—for a conception of the good and for a sense of justice—and gives priority to her interests in developing and exercising these moral powers. A capacity for a conception of the good is a capacity to develop, choose, and perhaps reconsider and revise if need be a conception of one's good, of what is valuable and choiceworthy in human life. A capacity for a sense of justice is the capacity to be disposed to play fair, to cooperate with others on terms regarded as fair.

In section 15, Rawls states that "the primary social goods, to give them in broad categories, are rights, liberties, and opportunities, and income and wealth." He adds that having a sense of self-respect or a sense of one's own worth is a very important primary good; the corresponding primary social good would be the social bases of self-respect. Section #29 stipulates that self-respect consists in a confidence that one's plan of life is worthwhile and also that one is competent to carry it out.

The idea that the principles of justice register the condition of individuals in terms of their holdings of primary social goods rather than in terms of their happiness, or the extent to which their life plans are fulfilled, or the extent to which their desires are satisfied, or the like, embodies a division of responsibility between the individual and society. Society (all of us regarded collectively) takes responsibility for providing all individuals a fair access to opportunities and resources, the primary social goods. What people do with their resources or opportunities for resources is their business, not the proper business of society or the government as society's agent. Given a fair social arrangement, each person is responsible for living her life as she chooses and for the resultant quality of her life. I may make a mess of my life, form the wrong friendships, choose the wrong job, marry badly or make bad romantic choices, but according to Rawls these matters are matters of individual responsibility not concerns of social justice. Choice of primary social goods as the measure of interpersonal comparisons involves a division of responsibility between individual and society and in this respect constitutes a major disagreement in principle between the Rawlsian and the utilitarian. On p. 80, Rawls writes: "It may be objected that expectations should not be defined as an index of primary goods anyway but rather as the satisfactions to be expected when plans are executed using these goods. . . . Justice as fairness, however, takes a different view. For it does not look beyond the use which persons make of the rights and

opportunities available to them in order to measure, much less to maximize, the satisfactions they achieve. 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. . . . Everyone is assured an equal liberty to pursue whatever plan of life he pleases as long as it does not violate what justice demands.” The primary goods idea is an aspect of Rawls’s commitment to the priority of right over the good. On a social contract view of justice, the members of society are not assumed to share ideals of the good life. They worship different gods, value different things. The basis for social contract agreement will not be agreement in detail about the substance of what is valuable and worthwhile in human life. We have to agree to disagree about these matters. Hence something like the primary social goods idea is central to Rawls’s theory of justice.

[[There are several primary social goods, so how do we decide, in general, how much of primary goods overall one has? Rawls speaks about an index of primary goods. This is a rule that enables one to determine, given anyone’s holdings of individual primary goods, what amount of primary goods overall one has. But Rawls never provides an index, and I do not see how he could provide one except by ranking the relative worth of different goods in a way he sometimes denies can be done in a diverse modern democracy. Rawls does give priority to the primary good of basic liberty in the Equal Liberty principle (but how do we decide, given there are several basic liberties, how to trade off one against others?). Absent an index of primary social goods, we don't really have a theory of justice from Rawls that meets his own standard.]]

[[Suppose we relax Rawls’s assumptions in the direction of realism by allowing that people differ in practical reasoning ability, including the ability to use primary social goods to fashion a good life. With equal primary social goods, Arneson chooses heroin and lives badly and you choose rock music and live well. But if Arneson’s bad life just reflects the bad luck that he was born with low potential for practical reasoning, why is the upshot fair, not a concern for social justice, by Rawls’s own lights? Another worry about the primary goods standard, voiced by Amartya Sen. Given that people differ enormously in physical traits, they vary in their ability to transform primary goods into what they might care about. One person has no legs, has to use all his primary goods to buy crutches, while a legless person can use her primary social goods to have a great life. Having equal primary social goods, the huge man must spend most of them to be adequately nourished, while the very small and thin-boned person can spend little to be adequately nourished and has lots left over to spend on other things. Again, why is this fair?]]

5. Strict compliance. The parties in the original position choose principles under this assumption: all persons will be disposed to comply with whatever principles are chosen. (This is entailed by finality, above.) This functions as a constraint on what may be chosen. Given human psychology, those in the original position are barred from choosing principles they know they likely could not steadily obey.

Section 26, the reasoning leading to the two principles of justice.

Given the set-up of the original position, Rawls argues that (1) the parties in the original position will choose according to the maximin decision rule and (2) maximin, the parties will then choose the two principles of justice advanced by Rawls in chapter 2 (Equal Liberty plus the combination of EFO, and the difference principle, with Equal Liberty taking strict lexical priority over EFO and EFO taking strict lexical priority over the difference principle).

Maximin.

The maximin decision rule says one should choose among alternative policies/acts by ignoring everything except the worst possible outcomes that might arise from any choice and selecting the policy with the best worst outcome.

Illustration of maximin. Suppose one is choosing a vacation and has three options: go to the beach, go to the mountains, or stay home. There are three possible states of the world that might arise after choice, and the value of the option chosen varies depending on the state of the world that ensue, as follows.

States of the world:	Options:		
	Go to beach	Go to mts.	Stay home
rain, no earthquake	5	10	-5
sunshine, no earthquake	20	15	-5
earthquake whether rain or sunshine	-200	-200	-100

For each option that might be chosen, the worst outcome is what happens if there is an earthquake. Maximin says one should ignore all other possible results and choose the option that yields the best payoff under this worst scenario. Maximin here says the correct choice is “stay home.”

An alternate decision rule: Maximize expected benefit (utility). This rule can be applied when one knows what are the possible outcomes of any choice one might make, and the value to one of that outcome if it occurs. In addition, one must be able to associate with each possible outcome of choice the probability that it will occur given one makes that choice. Maximize expected benefit says—one ought to choose by multiplying the value of each possible outcome of each possible choice by the probability it will occur if one makes the related choice. Sum the results for each choice. One ought to pick the choice that yields the highest sum.

Illustration of applying the maximize expected benefit rule: Suppose there is a .099 probability of rain, a .9 probability of sunshine, and a .001 probability of earthquake. Then the expected benefit of going to the beach is 5 times .099 plus .9 times 20 plus .001 times –200. This equals 18.295. The expected benefit of going to the mountains is 14.29 and the expected benefit of staying home is –5.095. The maximize expected benefit rule here says the correct choice is “go to the beach.” (If the measure of benefits is utility, then the rule becomes maximize-expected-utility.)

Why Maximin? As Rawls notes, maximin is in general not a good decision rule to follow. For one thing, it is insensitive to probability information. In the vacation decision problem above, maximin would say one ought to stay home even if the chance that an earthquake will occur is only one in a billion billion. For another thing, it tells us to ignore all the better than worst outcomes in deciding what to do, and this seems an extremely pessimistic or conservative way of deciding how to live.

Rawls notes that there are three circumstances such that the more it is the case that these circumstances obtain, the more maximin becomes a sensible decision rule. Rawls then adds that these three circumstances obtain in the original position, so maximin is a good decision rule to follow in the original position. The three circumstances are:

1. Knowledge of probabilities is unavailable.
2. We care very little if at all for the benefits we could gain above the level that maximin would yield us. (Suppose the benefits are dollars, maximin guarantees at least a million dollars in all situations one will face, and one cares nothing for money above a million dollars.)
3. Some of the outcomes which the policies not selected by maximin might yield would be unacceptable, disastrous. (In the vacation problem above, suppose the –200 payoff is a horrible intolerable disaster.)

[[The three circumstances do render maximin rational. But do they obtain in the original position? Take 2. In ordinary life, people rationally make risky life choices—they venture all their wealth on risky business strategies that might strike it rich; they choose sometimes to climb mountains by a

fun but risky route rather than take the absolutely safe but boring slogging route to the top; they drive across town, risking death on the highway, to get great bread at the great bakery across town, when they could play it safe and get stale bread from the grocery store within walking distance. And anyway, in the situation of the original position, we don't know what level of benefits maximin will yield or other decision rules might yield. We might find ourselves after the veil of ignorance is lifted in the ancient world, where maximin yields just a bit over bare subsistence, with lots of susceptibility to dread disease.

[[Take condition 3. By itself, even if accepted, this does not rationalize choice of maximin. Rather if we know some range of possible outcomes some decision rules could yield would be unacceptably disastrous, we should insist on a rule that gives priority to getting everyone to the threshold of acceptable outcomes. (The rule would be "make sure one has enough" not "make the worst off outcome as good as possible".) If we had knowledge of probabilities, and faced condition 3, we might choose a disaster avoidance rule: Choose whatever policy minimizes the risk of yielding a disastrous outcome.

[[Consider condition 1, we have no knowledge of probabilities. If this condition holds, this will be so because Rawls has set up the original position to make it so. The question then arises, why is complete ignorance of probabilities an appropriate condition to stipulate, when the task is to select principles of justice? Why not judge that Rawls is unfairly rigging the choice set-up so that the principles he favors emerges?

[[In the general conditions of human life, knowledge of probabilities is relevant information when deciding what one morally ought to do. Suppose I have a ship, and receive radio messages indicating I could steer left and maybe save 100 people or steer right and maybe save one person. Which choice should I make? It is relevant here, what the chances are that I would save any lives if I turned in either direction. Suppose there is a .99 probability that I will save one life if I steer right and a one in ten zillion chance that I will save any lives if I turn left toward the larger group in peril. This information affects what it is moral to choose. The intended effect of excluding probability information from the original position is to block choice of utilitarianism or any principle in the utilitarian family (e.g., sufficientarianism or prioritarianism). This is OK only if there are independent reasons why utilitarianism (or another principle in the family) is incorrect.]]

Rawls has another point concerning maximin and his principles. He says that maximin would not be a sensible rule in general, but it happens to give acceptable results for the expected range of decisions that we face when we face problems of social justice.

Finally, Rawls urges that if it is rational to maximin in the original position, it is rational to choose his principles. First we choose the Rawlsian general conception of justice—justice is maximin primary goods shares over the long run. In the special circumstances of modern society, which especially involve a level of social development that makes for a sufficiently wealthy economy, the special conception of justice becomes salient. This involves the choice of Rawls's two principles as described in chapter 2.

Section 27, the reasoning leading to the principle of average utility.

Rawls thinks utilitarianism is his chief rival. A version of utilitarianism (one ought to choose policies and institutional practices that maximize utility per person over the long run) might appeal to those in the original position or a quite similar choice situation. (If it is stipulated that parties in the original position are choosing to secure primary goods, then the contending principle would be, maximize the average level of primary goods over the long run.)

The idea behind choice of average utilitarianism in a version of the original position runs as follows. The individual is trying to do as well for herself as she can by choice of principles. But knowing no particular facts about myself, I don't know which of all the various people in society is going to be me. The veil of ignorance imposes a certain kind of impartiality on choice. Given the veil of ignorance, I cannot choose in a way that is partial to myself. But if I am self-interested,

then I want to maximize my own utility, and if I do not know anything about myself, the best I can do is to choose the principle that the average level of utility in society is to be maximized. If there are going to be three people in society, and I will be one, but I don't know which one, and I want to do the best I can to maximize my own utility, the best I can do is choose the principle that maximizes the average utility of the three.

Section, 28, some difficulties with the average principle.

1. P. 145: "There seem to be no objective grounds in the initial situation for assuming that one has an equal chance of turning out to be anybody." Without this assumption in place, the reasoning for choice of average utilitarianism collapses.

2. P. 152: "the utilitarian argument assumes that the parties have no definite character or will, that they are not persons with determinate final interests, or a particular conception of their good, that they are designed to protect." Utilitarian reasoning in the original position conceives persons as bare persons, who care only about maximizing the degree of satisfaction of whatever desires they might turn out to have.

Section 29, some main grounds for the two principles of justice.

Here Rawls appeals to the finality and publicity conditions to argue for his maximin conception of justice as superior to utilitarianism.

Recall that finality implies that one is constrained to choose only principles that one knows one will be disposed to support and obey once the veil of ignorance is lifted and one is living under the rule of those principles chosen.

Rawls urges that choice of utilitarianism in the original position amounts to gambling with one's fundamental life prospects. One takes a chance of becoming an aristocrat who gets cakes and ale and accepts the risk of becoming a peasant who gets potatoes and water. But once the veil of ignorance is lifted and one finds, say, that one is in the peasant position, knowledge that society conforms to utilitarian principle and that one's misery is offset by the greater good others enjoy will rankle. The strains of commitment given choice of utilitarianism might be too great. Since one cannot accept in the original position a principle one may not be able to obey when the principle is governing the society one inhabits, one could not accept utilitarianism in the original position. In contrast, maximin provides maximum assurance to the unluckiest in society, those whose initial life prospects are worst. So those with greatest reason to reject the rules of the game established in society will have least reason to do so if maximin is adopted. So, maximin should be adopted.

Appeal to the publicity constraint strengthens the appeal to finality. You might have a society ruled by a conception of justice that would inspire dissent and disobedience on the part of some members of society if they were aware that society was being run on this conception. But maybe we could solve this problem by keeping people from awareness of the basic constitution of society.

Example: Imagine a colonial society. The colonial administrators administer society according to a certain set of secular principles. But that society is governed according to these principles is a secret. What is announced and promulgated to ordinary members of society is that society is being ruled according to the dictates of the goddess Sheila and Sheila will cut off your private parts if you disobey. Valuing continued close association with their private parts, or perhaps loving the goddess Sheila, none of the commoners in society is inclined to disobey its rules. The society functions smoothly.

But this solution to the problem of securing compliance with basic rules of society and securing the stability of the basic structure of society over time is NOT AVAILABLE. It violates publicity. Publicity requires that the principles chosen will be a public conception of justice, available to be known by all members of society. The basic constitution of society is as it were engraved on the courthouse steps. Anyone can read and understand the basic constitution of society. But in the colonial society with Sheila worship just described, the principles that actually govern society are kept secret from the members.

If a utilitarian society under certain circumstances could be stable, attract the continued support and obedience of all members, only if it violates publicity, then Rawls says utilitarianism could not be chosen in the original position.

Rawls associates publicity with a Kantian ideal of respect for persons, to which utilitarianism is said not to be committed. Page 156: “the principles of justice manifest in the basic structure of society men’s desire to treat one another not as means only but as ends in themselves.” This is so because we agree to treat each other, in the basic structure of society, only by principles that would be chosen by all rational persons under conditions that are fair—the original position.

Insistence on finality and publicity also manifests the aspiration to treat people not merely as means but as ends in themselves. We agree to run society according to principles that all rational persons can accept and whose basic rules none would rationally violate. We agree also to run society by principles that are known by all or at least open to all so all can know these principles. In contrast, the utilitarian society might impose misery on some for the sake of greater gains for others, even if no one could rationally agree to submit to such misery (we fry the people of El Cajon on a hot griddle so that people in Marin County have a great life with hot tubs and air conditioning.) Also, nothing in utilitarianism carries any commitment to publicity. The imaginary colonial society might be one in which aggregate utility is maximized and if so, that is sufficient to justify choice of that type of society according to the utilitarian.

Rawls puts this point in another way: The society run on maximin lines caters more reliably to each person’s interest in the conditions that preserve her self-respect than would a society run on utilitarian lines. The utilitarian principle, according to Rawls, carries the message to all, if making you cannon fodder would maximize utility, that is what society would and should do.

[[Taken literally, insistence on finality seems to rule out any choice of principles, if the laws of psychology tell us there will be some people in any form of society who are strongly disposed to violate its rules and principles. But if we relax finality, as we must, then maybe utilitarianism can meet the condition. And anyway, the utilitarian will wonder why finality should be such a big deal. Suppose we can organize society so that people are in a state of bliss but a few of these blissful people will be disposed to disobey the basic rules; if this is the utility-maximizing choice, why not go for it?

[[Regarding publicity, and more broadly liberal conceptions of rights: for the utilitarian these will be justified, when they are, as means for maximizing utility in given circumstances. If publicity and democracy and entrenchment of various individual rights did not secure good quality lives for people, we should not favor them.

[[Regarding treating people not as mere means: the utilitarian accepts this formula and interprets it as follows: treating people not as mere means is treating them according to principles they could qua rational accept. Which principles are those? Whichever principles are morally correct, best supported by reasons. If utilitarianism is morally correct, best supported by reasons, then we treat people not as mere means by treating them according to the utilitarian principle.]]

Section 30, classical utilitarianism, impartiality, and benevolence.

According to Rawls it is an interesting fact that whereas average utilitarianism is a contender in the original position, aggregate utilitarianism is not. These two principles seem very similar but their deep structure is very different, as the social contract approach reveals. Classical (aggregate) utilitarianism can be modeled as the principles that would be the appropriate guide to choice of an impartial benevolent spectator who experiences everyone’s pleasures and pains and cares equally about all who do or might have such experiences.

Final comment. [[Notice that many notions of justice would not be considered at all in the original position as set up by Rawls. Consider the view that society should be set so that people gain happiness or good fortune in proportion to their virtue or deservingness. Or consider the view that society should be set to maximize (or maximin) human perfection, the achievement of true excellence. And so on. The original position, designed to block consideration of many principles, is acceptable only if there are good reasons for this blocking, for rejecting the principles not considered.]]

Final Final comment. [[Rawls could be correct that utilitarianism is a wrong or implausible candidate theory of justice even if his original position argument does not rule out utilitarianism and even vindicates it. In the original position Rawls sets up the choice of principles of justice in this way—consider individuals trying to advance their own interests by choice of principles. Force them to be impartial by imposing a thick veil of ignorance on them. But self-interested maximizing constrained by impartiality, roughly, yields utilitarianism. Rawls struggles mightily to resist this conclusion by complicating the decision procedure—people are choosing under uncertainty, and one should not be gambling with fundamental prospects, but rather should be choosing conservatively, etc. But he is struggling against the wind. It might be that if utilitarianism is unacceptable, and the original position construed plausibly yields choice of utilitarianism, then we should conclude the original position is not a good procedure for choosing principles of justice.]]