

RAWLS'S A THEORY OF JUSTICE

I. Overview of our Discussion of Rawls

A. Initial Comments on:

1. Scope of the Theory
2. Methodology
3. The Theory of Justice
 - a) The Model
 - b) The Principles

B. Detailed discussion of:

1. Scope of the Theory
2. Methodology
3. The Theory of Justice
 - a) The Model
 - b) The Principles

II. Initial Comments on Aspects of Rawls's Theory

A. Scope of the Theory

1. Rawls's theory is not a theory of morality (p. 17).
 - a) Though this is often forgotten, it is important for two reasons:
 - (1) It means that it is no criticism of Rawls that an action is morally obligatory but his theory doesn't establish that one ought to perform it. The action might be required by a portion of morality other than the theory of justice.
 - (2) Furthermore, if other moral considerations can override considerations of justice—*pace* Rawls—then it is no criticism of Rawls that his theory entails that one ought to do what is morally impermissible. Rawls's theory only gives us a fragment of morality: justice. It may well be that given *only* considerations of justice, one ought to perform an action that is forbidden when all moral considerations are weighed.
 - b) Therefore, criticisms that are in the form of counterexamples to Rawls's theory must be in terms of what *justice* requires, not what morality, rationality or any other normative theory requires.
2. Rawls's theory is not (at least not obviously) a full theory of justice. It is a theory of *distributive* justice.
 - a) If there is more to justice than distributive justice (and if this "more" isn't reducible to distributive justice) then there are two further lessons we must draw—lessons that parallel the above ones about morality.

- (1) It is no criticism of Rawls's theory that justice requires an act which his theory does not require if that act is required by a portion of the theory of justice other than distributive justice.
- (2) It is no criticism of Rawls's theory that his theory requires an action that is forbidden by justice if the prohibition comes from another portion of the theory of justice which overrides (in the particular case, at least) considerations of distributive justice.
- b) Therefore, criticisms in the form of counterexamples to Rawls's theory must be in terms, not just of what *justice in general* requires, but of what *distributive justice* requires.
3. These first two restrictions are fair enough and serve only to demarcate the sort of theory Rawls is giving.
4. However, Rawls goes on to limit the scope of his theory by arguing that the theory of distributive justice applies only within a certain range of circumstances. These are a modification of Hume's "Circumstances of Justice".
 - a) Hume's Conditions:
 - (1) moderate scarcity
 - (2) rough equality of abilities
 - (3) limited altruism
 - (4) non-self-sufficiency (interdependence, or non-independence)
 - b) To this list, Rawls adds: limited knowledge and strength of will.
5. While restricting the scope of his theory to distributive justice serves only to inform us of what he is theorizing about, the limitation of the theory of distributive justice to the circumstances of justice is a substantive philosophical claim about what sorts of conditions raise problems that theories of distributive justice need to resolve. This sort of restriction requires justification. We will question Rawls's justification for limiting the scope of a theory of distributive justice.

B. Methodology

1. There are two strands of justification in *A Theory of Justice*: one foundational and one coherentist. It is unclear to many how they are to fit together.
 - a) *Rawls's Foundationalism*: The foundationalism in Rawls is common to contractarians, though his is complex in certain ways. For Rawls, actions are justified by being in accordance with correct rules of distributive justice. Rules of distributive justice are justified by being in accordance with correct principles of distributive justice. Principles of distributive justice are justified by being shown to be rationally preferable for everyone from a certain perspective. A clearly hierarchical mode of justification is presented here.
 - b) *Rawls's Coherentism*: Rawls says that a theory is to be tested by its ability to accord with and explain our considered moral convictions in "reflective equilibrium." This is a dynamic process (p. 20) that is not at all foundational. Everything is up for revision in light of everything else. Principles can be changed in light of our

considered conviction about the rightness of actions and the very conditions under which the principles are to be chosen (a matter that is at the very foundation of Rawls's contractarianism) can be changed if they lead to particular judgments that we are convinced are wrong.

2. *Fitting the foundationalism together with the coherentism:* My way of understanding the way these two modes of justification fit together in Rawls is based on distinguishing between justification *within* a theory and justification *of* a theory. I take Rawls to be giving a coherentist (reflective equilibrium) argument for a theory that is, in its internal structure, foundational.
- C. *The Theory:* Rawls's theory has two parts: what he calls a "model" and his principles of justice.
1. Rawls's Model of Justice
 - a) *The Idea of a Model of Justice:* Rawls never gives a very clear account of what a "model" of justice is. Here is a stab: a model of justice is (or yields) a decision procedure for determining the justice of rules, principles and actions (*etc.*) which is not only effective, but explanatory.
 - (1) *Effectiveness:* A decision procedure is effective if it correctly sorts items on the basis it is supposed to sort them.
 - (2) *Explanatory:* A decision procedure may be effective without being explanatory. If my grandmother is inerrant with respect to matters of justice and has passed judgment on the justice of everything, then the decision procedure "something is just if, and only if, Hubin's grandmother believes it is" is effective. It is not, however, explanatory at all. Just what explanatoriness consists in is a large matter that I won't try to address here.
 - b) Elements of Rawls's Model of Justice
 - (1) *Early Intuition:* In "Outline of a Decision Procedure for Ethics" (*Philosophical Review*, 1957), Rawls takes the principles of justice to be those principles that describe what would be agreed to by all who impartially consider the issues. This could lead either of two developed views:
 - (a) *Ideal Observer Theory:* One imagines the impartial party considering the issues to be fully informed and to be impartially benevolent—to be equally concerned with each person's well-being.
 - (b) *Hypothetical Contractarianism:* One imagines the principles of justice to be those that would be chosen by individuals who care only to improve their own position but are constrained in certain ways from doing so. This, Rawls claims, can achieve the result of impartial benevolence without imagining impartial benevolence on anyone's part.
 - (2) *Later Intuition:* For various reasons that we will discuss later, Rawls prefers the hypothetical contractarian approach as a way to develop the above insight about the nature or principles of justice.
 - (a) The principles of justice are those that would be agreed to by people in a fair initial situation of choice.

- (i) *Illustrative Example: A modified version of Risk.*
 - (ii) *Characterization of the Initial Situation:* Rawls's interpretation of this initial situation is called "the original position".
 - [a] *Rationality:* Agents in the Rawlsian original position are assumed to be rational in the sense of taking efficient means to desired ends.
 - [b] *Mutual Disinterest:* Agents in the original position are assumed not to take an interest in the payoff of others nor in the differences between their own payoffs and those of others.
 - [c] *Veil of Ignorance:* Agents in the original position do not know particular facts about themselves. (That is, they do not know their own age, sex, race, religion, life goals, etc.)
 - [d] *(Implicit) Veil of Volition:* Agents in the original position are not motivated by their real life motivations except insofar as those are identical with the motivation for primary goods discussed below.
 - [e] *Desire for Primary Social Goods:* Agents in the original position are motivated to attain for themselves the highest quantity of primary goods possible. These are: rights & liberties, income & wealth, powers & opportunities, and the social bases of self-esteem.
 - c) *Justification of Rawls's Model of Justice:* Rawls has two tasks to complete in order to justify his model of justice.
 - (1) *The Problem of Content:* Rawls has to justify his characterization of the initial situation with respect to each of the above conditions. We could construct a hypothetical contractarianism that has different assumptions about the membership in the initial situation, the knowledge of the contractors, the motivation and dispositions of the contractors, etc..
 - (2) *The Methodological Problem:* Rawls has to show the relevance of agreement in the original position (or some other conception of the initial situation) to the justice of the principles.
2. Rawls's Principles of Justice:
- a) Statement of the two principles of justice.
 - (1) First Statement: p. 60 (p. 53, revised edition).
 - (2) Final Statement: p. 302 (p. 266, revised edition).
 - b) General Conception of Justice.
 - (1) First Statement: p. 62 (p. 54, revised edition).
 - (2) Final Statement: p. 303 (p. 266, revised edition).
 - c) Justification of the Principles of Justice

- (1) Rawls would *like* to show that his general conception of justice and the two principles that he believes follow from it are derivable from the assumptions of the original position. He believes this to be too ambitious.
- (2) Rawls settles for arguing that his conception and the principles of justice are preferable to the classical alternatives—especially utilitarianism.

III. The Scope of Justice

A. The Vagueness of the Concept of Justice and the Scope of Justice:

1. *Instances where considerations of justice seem clearly to arise:* The terms of cooperation between normal adult humans with conflicting interests who are engaged in an on-going cooperative scheme that produces social benefits.
2. *Instances where considerations of justice seem clearly not to arise:* The interactions between a virus and the host organism.
3. *Borderline cases:*
 - a) *Nonstandard relation:* relations between nonoverlapping generations.
 - b) *Nonstandard relata:* relations between young children and their parents.
 - c) *Nonstandard circumstances:* extreme scarcity, desert island.

B. Importance of Clarity on the Scope of Justice

1. *Theoretical:* If we identify the scope of justice incorrectly, we will be misguided in our criticisms and evaluation of a theory of justice.
2. *Practical:* If we identify the scope of justice too narrowly, we may give no guidance or the wrong guidance when evaluating the justice of some action or institution. If we identify it too broadly, we may give guidance where we shouldn't give any or give the wrong guidance in such evaluations.

C. Hume on the "Circumstances of Justice"

1. Rawls largely accepts Hume's treatment of the circumstances of justice (pp. 126-30, 109-112 in the revised edition).
 - a) It is crucial to his argument at several points that we will discuss later.
 - b) It has come to be seen by many as uncontroversial.
2. Hume's Conditions:
 - a) Moderate Scarcity
 - b) Limited Altruism
 - c) Rough Equality of Capacities and Aptitudes
 - d) Non-independence
3. Ambiguities in Hume's Account of the Circumstances of Justice: Hume holds a thesis about the relation between justice (in some sense) and these conditions. Just what the thesis is unclear because of two ambiguities. Hume is unclear both about the nature of the relationship and the nature of the things related: the relata.

- a) *Relata*: Are Hume's conditions related to the *concept* of justice or to *institutions* of justice?
 - b) *Relationship*: Is Hume asserting a *logical*, a *factual*, or a *normative* relationship between the conditions of justice and whatever they are supposed to be related to?
4. *Disentangling the Ambiguities*: Five distinct interpretations are suggested by things Hume says:

	Relata: Hume's Conditions and . . .	Relationship
Logical Interpretation	Concept of Justice	Conceptual (Logical)
Epistemic Interpretation	Concept of Justice	Factual (Causal)
Ontological Interpretation	Institutions of Justice	Factual (Causal)
Deontic Interpretation	Institutions of Justice	Normative
Utility Interpretation	Institutions of Justice	Factual if 'utility' = pleasure Normative if 'utility' = desirability

5. Evaluating the Interpretations

a) Logical Interpretation:

(1) *Account*: The conditions of justice are necessary (and perhaps sufficient) for the concept of justice to be applicable. Assertions about justice *and injustice* logically presuppose that the conditions obtain in the relevant situation.

(2) Textual Support:

(a) *A Treatise of Human Nature*, p. 496.

(b) *An Enquiry Concerning the Principles of Morals*, p. 23.

(3) Criticism:

(a) Contra Moderate Scarcity:

(i) Lifeboat Injustice

(ii) Distribution of "Sentimental Property" in Abundance

(b) Contra Limited Altruism

(i) Injustice in a "Society of Ruffians"

(c) Contra Rough Equality

(i) Torment of the Weak

b) Epistemic Interpretation:

- (1) *Account*: The conditions of justice are necessary for us to acquire the concept of justice.
 - (2) Textual Support:
 - (a) *A Treatise of Human Nature*, p. 495.
 - (b) *An Enquiry Concerning the Principles of Morals*, p. 16 (possibly) and p. 17.
 - (3) *Criticism*: There is little attraction to this interpretation if the logical interpretation is false. But even if that interpretation were correct, the motivation for the epistemic interpretation is unclear. If Hume can have the concept of the-concept-of-justice-not-applying-when-the-conditions-of-justice-do-not-hold, why should we think people not subject to these conditions incapable of forming the concept of justice by imagining the conditions of justice to obtain.
- c) Ontological Interpretation:
- (1) *Account*: The conditions of justice are necessary for institutions (actual social rules of justice in practice, *etc.*) to develop or persist. Hume seems to think that institutions of justice are only socially useful under his conditions (as we'll see later). This seems to play a role in his (apparent) acceptance of the ontological interpretation.
 - (2) Textual Support:
 - (a) *An Enquiry Concerning the Principles of Morals*, p.17 & 20.
 - (3) Criticism:
 - (a) *Existence and Utility*: It is not true that institutions can come into existence and/or persist only if they are socially useful. Thus, the connection to social utility, even if Hume establishes it, doesn't justify the ontological interpretation.
 - (b) *Existence and Effectiveness*: Hume seems to believe that institutions of justice could not be *effective* if the conditions of justice did not obtain (*Enquiry*, p. 23). And, for basic institutions, to be is to be effective. But, it is not true that each of the conditions of justice is necessary for effectiveness of institutions of justice. For example, violation of rough equality would not necessarily lead to the ineffectiveness of institutions of justice if individuals were not purely selfish.
- d) Deontic and Utility Interpretations:
- (1) *Account*:
 - (a) *Deontic*: The conditions of justice are necessary for the rules of justice to be morally obligatory.
 - (b) *Utility*: The conditions of justice are necessary for the observance of the rules of justice to be socially useful.

- (2) *Textual Support:* These claims are not conjectures or tenuous interpretations. They are central to Hume's strategy in the *Enquiry*.
- (a) *An Enquiry Concerning the Principles of Morals*, p. 20.
- (3) *Hume's Strategy:* Hume seeks to show that considerations of justice are grounded on considerations of utility by arguing that, where the circumstances of justice fail to obtain, justice is neither useful nor obligatory.
- (a) Criticism of the Strategy:
- (i) Even if Hume is right in all he claims about the implications of the violation of the circumstances of justice for the utility and moral obligatoriness of justice, it would not show that rules of justice are obligatory only when they are socially useful.
- [a] *Logic of the Argument:* Hume seeks to show by his examples that: $\sim CJ \supset (\sim O \ \& \ \sim U)$. ['CJ' means 'the circumstances of justice obtain'; 'O' means 'rules of justice are obligatory'; and 'U' means 'rules of justice are socially useful'.] He seems to conclude from this that: $\sim U \supset \sim O$ (or, equivalently, $O \supset U$)
- [i] Counterexample: This is left as an exercise for the interested reader.
- (i) Even we grant Hume the above implication (or just give him the conclusion that $\sim U \supset \sim O$, this only establishes an implication (logical or material depending on how we interpret ' \supset '). It does not establish an explanatory relationship: that the utility of rules of justice *explains* their moral obligatoriness.
- (4) Criticism of the Deontic and Utility Interpretations
- (b) *Shipwreck Example:* Adherence to the rules of justice under conditions of severe scarcity might be able to increase the number of people receiving the minimum amount to survive. This appears to make the rules both socially useful and morally obligatory.
6. *Summary of Hume on the Circumstances of Justice:* It is false that Hume's conditions are each individually necessary presuppositions of justice in any of the above interpretations. That is to say, no single interpretation renders this claim true of all the conditions. Still, some of these conditions may:
- a) *define the normal conditions of justice:* This means that our intuitions may be "honed" for such cases and we need to proceed carefully in considering other sorts of cases.
- b) *influence the nature of the rules of justice:* This would suggest that we might want to understand the rules of justice as having different implications for different situations.
- c) *be conditions of justice in one or more of the senses Hume suggests:* For example, the non-independence condition may be a logical presupposition and seems clearly to be an ontological one. (*Enquiry*, p. 24.)

- D. *Rawls's Interpretation of the Circumstances of Justice:* Rawls claims to add nothing essential to Hume's discussion. While it may be true that he adds nothing worthwhile, he does make significant additions.
1. Rawls's Conditions:
 - a) rough equality of physical and mental powers;
 - b) moderate scarcity;
 - c) coexistence together at the same time in a definite geographical region;
 - d) similar needs and interests;
 - e) self-interest (in some sense); and,
 - f) shortcomings of knowledge; thought and judgment.
 2. *Rawls's Conception of the Relation between these Conditions and Justice:* Rawls's conception of the relation between his conditions and justice is different from Hume's. Rawls claims that these are the normal conditions under which human cooperation is both possible and necessary (p. 126).
 3. Criticisms of Rawls's on the Circumstances of Justice:
 - a) The Conditions:
 - (1) *Conditions #3 (c above):* It is not clear in what way geographical location is at all relevant to institutions and rules of justice (except in a wholly contingent and uninteresting sense). Coexistence at the same time seems to rule out, *by fiat*, justice between non-overlapping generations.
 - (2) *Condition #4 (d above):* No justification is given for this. Presumably Rawls wants #3 and #4 to work together to ensure non-independence and, perhaps, a certain kind of conflict. But they do so poorly.
 - (3) *Condition #6 (f above):* It is not clear what relevance our shortcomings of knowledge, thought and judgment have on institutions of justice.
 - b) The Relation of these Conditions to the Possibility and Necessity of Human Cooperation:
 - (1) The Possibility of Human Cooperation:
 - (a) We have already argued that the ontological presupposition interpretation suggested by Hume is not correct. If not, then institutions of justice can exist when these conditions don't obtain. But institutions of justice just *are* schemes of social cooperation.
 - (b) Extreme scarcity doesn't make human cooperation impossible, nor does utter selfishness, great inequality of abilities, living arbitrarily far away from one another, *etc.*
 - (2) The Necessity of Human Cooperation:
 - (c) *The Interpretation of 'Necessity':* Rawls must mean something like 'necessary for humans to accomplish other goals' or 'maximize their utility'

or some such thing because, of course, human cooperation is clearly not rendered *necessary* in any strict sense by the conditions he lists.

- (i) The view that, in the absence of Rawls's conditions, human cooperation is not necessary to accomplish our ends seems to assume a view that has been called "possessive individualism": that the aims of humans are solely to secure their material well-being. If humans have intrinsic goals of being involved with others (playing games, talking, loving, *etc.*), then human cooperation is necessary to achieve our ends regardless of the abundance of external wealth.
- (ii) If we restrict our attention to cooperation for the production of material goods, then true Humean abundance of desired items (in their consumable form) would seem to render human cooperation unnecessary. But this seems to be the only one of Rawls's conditions that has this effect.

E. *Implications of a Broader Conception of the Scope of Justice*: These criticisms of Hume and Rawls suggest that the concept of justice has broader applicability than either seems to suppose. This has some important implications. If it is true, then:

1. Rawls's theory is subject to more stringent tests. In particular, Rawls will have to show that his theory leads to acceptable conclusions across a wider range of cases than he supposes.
 - a) I shall argue later that his principles of justice lead to clearly unacceptable consequences when applied to the problem of justly distributing extremely scarce necessary resources.
2. Rawls cannot, in deriving his principles, employ knowledge that the conditions of justice (Hume's or his own) obtain.
 - a) Interestingly, in a way this will help Rawls out with the first problem, I think. Once this assumption is dropped, his model of justice will not lead to the principles Rawls thinks it will. Arguably, it will lead to more plausible principles. If so, this is a happy result for his model, if not for his principles.
3. Independently of Rawls's theory, the difficulty of stating and defending principles of justice is increased.
4. *Possible Payoff*: Once we get clearer about the real conceptual presuppositions of justice, we may be in a better position to solve the hypothetical contractarian's methodological problem (see p. 4 above and the section below on the justification of hypothetical agreement). I will argue that with a better understanding of what *is* presupposed by the concept of justice, we will see better why it is relevant to justice what rational agents would *agree* to in some nonactual situation.

IV. The Justification of Hypothetical Agreement

A. The Role of the Original Position

1. Conceptions of the Relation between Morality and Rationality (Reasons for Acting): There have been various attempts to establish a connection between morality and rationality—to show in what sense it is rational to act morally.

- a) *Morality as Subjective Rationality*: One might understand morality to be simply what is subjectively rational in the sense that it maximizes the agent's subjective expected utility.
 - (1) This provides a ready answer to the question, "Why should I be moral?" Acting morally is just doing what one has a reason to do in the standard economic sense of rationality.
 - (2) This account doesn't accord with our conception of morality for at least two reasons.
 - (a) It allows morally irrelevant features to count.
 - (i) One respect in which it does this is that it allows allow distinctions between individuals based on nothing more than who they are.
 - (b) It leads to an extreme form of relativism.
 - (3) Some deny that a theory with these implications will even do as an account of rationality.
 - (a) Kantian Intuition: Spock's Irrationality.
 - (b) Contrary Intuition: The Judge's Ass.
 - b) *Morality as Impartial Rationality*: One might understand morality as impartial rationality. Such rationality could not distinguish cases that differed merely *in personae*.
 - (1) This is commonly held to be a necessary feature of morality. Rawls requires this both by a formal constraint and, implicitly, by the veil of ignorance.
 - (2) Some (Hare, perhaps) see impartial rationality as sufficient for a moral judgment. This puts no necessary substantive limits on the grounds of moral judgment.
 - (a) Problem of the Fanatic: Heliogabalus.
 - c) *Morality as Perspectival Impartial Rationality (Perspectival Rationality, for short)*: One might understand moral reasons to be reasons in the following sense. From the perspective of morality, these are considerations that determine how it is rational to act.
 - (1) *Modesty of the Position*: This position doesn't attempt to show that moral reasons are rationally binding on all people—*i.e.*, that they are reasons from every person's evaluative perspective. Furthermore, it doesn't attempt to show that moral reasons are reasons for everyone insofar as they choose impartially.
2. *Rational Justification of the Principles of Justice and Choiceworthiness in the Original Position*: The imposition of the veil of ignorance indicates that Rawls doesn't hold that the principles of justice a to be identified with or derived from the principles of subjective rationality. Rawls's acceptance of the motivational assumptions indicates that he doesn't think that impartial rationality implies the principles of justice. Rather, he seems to believe that the principles of justice are rational in the sense that they would be chosen by subjectively rational individuals who were precluded from acting partially and were pursuing certain specified ends.

- a) Unless everyone necessarily adopts the perspective in question (pursuit of primary goods), reasons derived from the principles of justice will not necessarily be reasons for every person *in the sense that they are rationally binding on everyone*.

- (1) Rawls never argues that everyone adopts the perspective of the original position.

- b) The point of showing that the principles of justice are rationally acceptable from the point of view of the original position is not primarily motivational. Nor is it to show that insofar as we are rational we will act justly. Rather, it is justificatory. Rawls seeks to justify the principles of justice from “the moral point of view”—to show that they are rationally acceptable from the point of view morality dictates for choosing such principles.

B. The Role of Hypothetical Agreement

1. *Rawls's Views*: Rawls seems to be of two minds concerning the role of hypothetical agreement.

- a) *Heuristic Device*: The fact of hypothetical agreement plays no justificatory role with respect to the principles of justice. Rawls says, for example: “One should not be misled, then, by the somewhat unusual conditions which characterize the original position. The idea here is simply to make vivid to ourselves the restrictions that it seems reasonable to impose on arguments for principles of justice, and therefore on these principles themselves” (p. 16). And, “[o]ne way to look at the idea of the original position, therefore, is to see it as an expository device which sums up the meaning of these conditions and helps us to extract their consequences” (p. 19).

- b) *Taking Hypothetical Agreement Seriously*: The fact of hypothetical agreement plays an important justificatory role with respect to the principles of justice. Rawls says: “The merit of the contract terminology is that it conveys the idea that principles of justice may be conceived as principles that would be chosen by rational persons, and that *in this way conceptions of justice may be explained and justified*” (p. 14-15, emphasis added).

2. *Dworkin's Challenge*: Hypothetical agreement can only be a heuristic device. It can serve no essential role in justifying the claim that Rawls's principles are the principles of justice. At most it can point to factors that provide such justification. Dworkin admits a necessary extensional equivalence between the principles of justice and what people would agree to in the original position but denies that facts about hypothetical agreement play any role in the justification of the principles of justice. His criticism holds that hypothetical agreement in the original position is a “theoretical dangler” in the justification.

- a) Arguments:

- (1) “A hypothetical contract is not simply a pale form of an actual contract; it is no contract at all” (*Taking Rights Seriously*, p. 151).

- (a) *Criticism*: This would be a good point to make if anyone had ever been confused about it. Rawls should admit that the force of hypothetical agreement doesn't come from the notion of a *contract*. That notion *is* a mere heuristic. But, Rawls should say that it is a heuristic for the essential

justificatory claim that, from a particular (and privileged) point of view, people would agree to his two principles.

- (2) Dworkin's Poker Example (*TRS*, p. 151).
 - (3) Euthyphro Argument.
- b) Dworkin's Concession:
- (1) Parentalistic (Paternalistic) Intervention (*TRS*, p. 152).
 - (2) *How the Concession Undermines his Position:* While Dworkin attempts to distinguish the case of parentalistic (paternalistic) intervention from that of justification of the principles of justice, his distinctions are not real. The real distinctions that exist, do not justify his acceptance of hypothetical agreement for the parentalistic (paternalistic) intervention together with rejection of it for the case of justifying principles of justice.
 - (a) Dworkin's Proposed Distinction:
 - (i) Temporal Asymmetry:
 - [a] There is no temporal asymmetry between the two cases. In neither case is the attempt to show that an earlier agreement is binding on one's later self.
 - (ii) Different Circumstances:
 - [a] There is no asymmetry between the two cases in this respect either. In both cases one asks what a person would agree to in one set of circumstances to determine the justification of applying principles with respect to him or her in other circumstances.
 - (b) A Real Distinction:
 - (i) Parentalistic (paternalistic) justification seeks to justify intervention from a suitably corrected point of view of the person interfered with.
 - (ii) Justification of the principles of justice takes place from the moral point of view (or, more narrowly, from a point of view that is morally appropriate for choosing principles of justice).
 - (iii) This distinction is real, but it doesn't support Dworkin's charge that hypothetical agreement is morally irrelevant in the employment Rawls makes of it—especially given his admission that it is morally relevant in the case of parentalistic interference.
 - (3) Dworkin's argument against Rawls's use of hypothetical agreement is undermined (at least to the degree that arguments are undermined by *ad hominem* rejoinders) by the following consideration. His arguments against Rawls's employment of hypothetical agreement would, if sound, also impugn the use of hypothetical agreement in the case of the parentalistic interference.
 - (a) Hypothetical agreement to parentalistic interference is not some pale form of agreement.

- (b) If you can give an argument that *A* would agree to such treatment, then it must be in virtue of some properties of the treatment together with some facts about *A*. (Dworkin ignores the issue of the nature of the hypothetical agreeer, but it must be assumed.) Therefore, by Dworkin's reasoning, hypothetical agreement *even in the parentalistic case* is unnecessary to justify action.
 - c) *Conclusion of Discussion of Dworkin's Challenge*: Dworkin's arguments against the moral significance of hypothetical agreement are not conclusive and his position is inconsistent.
- 3. The Role of Hypothetical Agreement:
 - a) The hypothetical agreement which is at the core of modern hypothetical social contract theories reflects certain presuppositions of justice. Problems of justice are seen as arising between individuals who are both moral patients and moral agents.
 - (1) *Moral Patients* are individuals the treatment of which is of direct moral significance simply because of how it affects them. Presumably, such characteristics as having interests, desires and concerns and being the subject of experiences like pleasure and pain are relevant to one's status as a moral patient. The social contract theory reflects this aspect of persons by presupposing that the contractors have interests that may come into conflict.
 - (2) *Moral Agents* are individuals whose actions it is appropriate to evaluate morally in such deontic categories as right or wrong, permissible or impermissible, praiseworthy or blameworthy, *etc.*. Presumably such individuals must be capable of choice that is rational and free in some morally significant sense. The social contract theory reflects this aspect of persons in two ways:
 - (a) It presupposes agents who are capable of contracting (agreeing).
 - (b) It presupposes agents who are capable of acting on the terms of the agreement.
 - b) Thus, the model of hypothetical social contract reflects the presuppositions of the concept of justice (the true "scope of justice"): that problems of justice are those that arise when moral persons (those who are both moral agents and moral patients) are in situations where they have potential conflicts of interests that can be ameliorated by cooperation but where there are also conflicts of interest about the terms of mutually beneficial cooperation.
 - c) Analogy to Filial Duties.
- 4. Contrast with a Utilitarian *Model of Justice*
 - a) Such a model views the benefactor/beneficiary relationship as the only directly morally relevant relationship. Each person is to view himself as a utility spigot and everyone (himself included) as a utility receptacle. The moral decision is made as if the actor is the only moral agent and everyone is a moral patient. To the extent that others are relevant in our moral deliberations, they are relevant only as moral patients.

- b) The social contract model, in contrast, recognizes the existence of a moral *community* within the moral universe—a moral community consisting of moral persons—and it exploits this to explain what is common to the diverse duties of justice.
 - c) Dworkin's conception of the role of hypothetical agreement destroys this important distinction between social contract theories and standard utilitarianism, even if it maintains the substantive requirements of Rawls's theory.
 - (1) Instead of asking us to maximize social utility, Dworkin would ask us to act so as to promote people's "antecedent interests". But, like the utilitarian model, this asks us to consider others only *qua* moral patients.
 - d) This isn't a point about the content of the utilitarian principle but about the mode of justification it is given. John Harsanyi accepts a utilitarian standard of justice but grounds this in a social contract theory. He argues that people in a properly constructed initial situation would agree to a utilitarian principle of justice. Such a view preserves the contractarian model. If the argument is sound, it gives the sort of justification to utilitarian principles that Rawls claims for his two principles of justice.
5. The notion of the moral community being a presupposition of justice and of the importance of showing that the principles of justice are not only morally binding but binding *qua* principles of justice helps answer some questions that have been posed to the social contract theorist:
- a) Why is there more than one person in the Original Position?
 - (1) Group membership is required not to get the substantive principles (which would be the same even if there were only one person) but to explain why:
 - (d) the duties of justice are *owed to others*; and,
 - (e) others have a *right against us* to their performance.
 - b) How is the social contract model preferable to an ideal observer theory?
 - (1) The social contract model the intuitions just mentioned in a way what cannot be captured in the ideal observer theory.

V. The Principles of Justice

A. Statement of the two principles of justice.

1. First Statement: p. 60 (p. 53, revised edition).
2. Final Statement: p. 302 (p. 266, revised edition).

B. General Conception of Justice.

1. First Statement: p. 62 (p. 54, revised edition).
2. Final Statement: p. 303 (p. 266, revised edition).

C. Reasoning for the Two Principles of Justice

1. Formal Constraints on the Concept of Right

- a) *The Role of Formal Constraints*: Since the principles of justice are a subset of the principles of right action, if there are any formal constraints on the concept of right action, these will apply as well to principles of justice. In this section, Rawls claims that there are such constraints and attempts to describe and defend them. The constraints are summarized as follows: “a conception of right is a set of principles, general in form and universal in application, that is to be publicly recognized as a final court of appeal for ordering the conflicting claims of moral persons” (p. 135, revised edition p. 117).
- b) *The Constraints*:
- (1) *Generality*: Principles of right action must contain no proper names or “rigged definite descriptions”. (No account of a “rigged definite description” is given.) They must be formulable in general—in principle repeatable—properties.
 - (2) *Universality*:
 - (a) Fundamental principles of right action must apply to everyone; they must apply solely in virtue of one’s being a moral person. Rawls seems to think that this implies that everyone can understand the principles and use them in deliberation, though *this* requirement seems to be more than just universal applicability.
 - (b) Principles of right action cannot be such that it is self-contradictory or self-defeating for everyone to act on them.
 - (c) Principles of right action cannot be such that it is reasonable to act on them only if others do not.
 - (d) [Note: These last two requirements lead Rawls to claim that principles of right action “are to be chosen in view of the consequences of everyone’s complying with them. This can’t be a formal constraint on the concept of right. Presumably he is claiming that the fact that the parties in the original position choose on the assumption of strict compliance, the original position *reflects* this formal constraint on the concept of right.]
 - (3) *Publicity*: Rawls doesn’t state this so that it appears to be a *formal* condition on the concept of right. Here is a stab: principles of right must be publicly acceptable; they may not be principles that must be kept secret. (I assume that the ‘must’ in this characterization is one of *moral* necessity—not merely logical or physical necessity. If so, this is closely related to the requirements 2 and 3 under ‘universality’.)
 - (4) *Ordering*: The principles of right (collectively) must impose an ordering on conflicting claims. The ordering should be complete and transitive. By the end of his discussion, Rawls strengthens the requirement to state that the ordering must be “based on certain relevant aspects of persons and their situation which are independent from their social position, or their capacity to intimidate and coerce” (p. 134, r.e. p. 116). [It is not plausible, I think, to understand this last requirement as a *formal* requirement.]
 - (5) *Finality*: The principles of right must be the “final court of appeal in practical reasoning” (p. 135, r.e. p. 116). These principles (considered collectively)

override any other practical principles. [Again, this doesn't seem to be a *formal* feature of the principles. It might be better to think of it as a *functional* requirement.]

c) Interpretations of the Constraints

(1) These putative formal constraints on the concept of right are embodied in the characterization of the original position. To the degree that we take them actually to be such constraints, they help to justify that characterization.

(a) *Textual support*: "The situation of the persons in the original position reflects certain constraints" (p. 130, r.e. p. 112).

(2) These putative formal constraints serve as a filter on what principles the persons in the original position can consider. They are not embodied in the characterization of the original position and they do not function to support that characterization. Rather, they function independently, prior to the original position argument for the principles of justice.

(a) *Textual support*: "The several kinds of egoism, then, do not appear on the list presented to the parties. They are eliminated by the formal constraints" (p. 136, r.e. p. 117).

d) Critique

(1) *Criticism of Individual Constraints*: While several of the constraints (generality and the first condition of universality) are widely accepted as relatively noncontroversial "conditions of adequacy" for a principle of right action, others are highly controversial. Utilitarians (Sidgwick, for example), among others, have thought that a correct principle of right could violate the publicity requirement. Many philosophers hold that finality is not a constraint on a principle of right. (Of course, the principles of *moral* rightness collectively are final concerning what one *morally* ought to do. But they need not be the final court of practical reason. That is, it may be *reasonable* for someone to act contrary to moral principles.)

(2) *Criticism of the Project*: If the first of the above two interpretations is correct (and if Rawls were right that these were widely-shared and weak moral assumptions), then this section is not especially troubling. He is simply doing what he has said he will do: justifying the characterization of the original position by appealing to weak and widely-held moral convictions. However, if the second interpretation is correct, then it is mysterious what this section is doing. It might make sense to slap some formal requirements on the principles of justice if that were necessary to ensure that the contractors not "go wrong". (It would diminish the elegance of the theory, perhaps.) However, the only principles that Rawls says are ruled out by the formal constraints are certain versions of egoism, and he specifically argues that these would not be chosen by the contractors anyway. So, what is the role of these constraints?

2. Characterization of the Original Position

a) Equality

- (1) The equality of the parties in the original position is emphasized in § 40 as part of the “Kantian Interpretation” of justice as fairness. The intuitive idea is that none are in a position to dominate others. Since the other restrictions ensure that every party in the original position will reason in exactly the same way to exactly the same conclusion, this assumption does no work in deriving the principles of justice from the assumptions of the original position. It appears to be included primarily so that the Kantian interpretation seems more plausible.
 - (a) It might have been better for Rawls to argue that the other assumptions of the original position work to ensure that the model respects the moral equality of people. Making a separate assumption of equality of people in the original position seems gratuitous.
- b) Freedom
 - (1) Again, this is stressed in § 40. It doesn't seem necessary for the derivation of the principles because all that needs to be shown for that is that the principles are rationally preferable to alternative principles. However, if I am right about the role of hypothetical agreement in justifying principles from the (corrected) point of view of the agent, it is crucial that each of us would have freely chosen these principles from the relevant point of view. If so, then the assumption of freedom plays a role in justification even though the principles could be derived without this assumption. It plays a role in showing that these principles are properly related to us as rational agents.
- c) Veil of Ignorance
 - (1) The veil of ignorance precludes the contractors from knowing particular facts about:
 - (a) themselves as individuals—their race, sex, ethnicity, religion, height, weight, sexual orientation, marital status, intelligence, your conception of the good, *etc.*,
 - (b) their societies—level of development, natural resources, place in time, power relative to other societies, relative size of various classes in society *etc.*,
 - (c) their role in society—economic status, political office, *etc.*.
 - (2) *Justification:* Rawls seeks to preclude from the contractors knowledge of facts that are both morally irrelevant and tend to lead to disagreement.
 - (a) As we will see later, not all of the knowledge screened out by the veil of ignorance is plausibly seen as morally irrelevant. (In particular, the knowledge of the size of the various classes in society is probably not morally irrelevant.)
- d) (Implicit) Veil of Volition
 - (1) Rawls seems to assume that precluding the contractors from *knowing* their real life conceptions of the good precludes them from being motivated by them. This is an overly cognitive (belief-based) conception of human motivation.

Clearly, Rawls wants to prevent the contractors from being motivated by their real life conception of the good, too.

e) Amoral motivation

- (1) Rawls assumes that the contractors are not motivated by moral considerations. The moral principles are supposed to come out of the original position. To assume moral motivation would be circular.

f) Rationality

- (1) Rawls assumes that the contractors are rational in the standard economic sense of taking efficient means to desired ends. This is a formal (rather than substantive) sense of 'rational'. In this sense, one could be quite rational even if one pursued the most bizarre ends. To be a bit more precise, Rawls assumes that each person has a coherent set of preferences over outcomes. This requires, among other things, that each person's preferences be:

- (a) Transitive and acyclic—if x is preferred to y and y to z , then x is preferred to z and it is not the case that z is preferred to x ;
- (b) Complete (universal)—that is, for *any* two outcomes, either the first is preferred to the second or the second is preferred to the first or the individual is indifferent between the two outcomes

The contractors are assumed to choose more preferred outcomes over less preferred outcomes.

- (2) *Non-Envy*: Rawls assumes that the contractors are not envious—they do not accept a loss for themselves so that others have less as well. He thinks this is a break with economic conceptions of rationality. However, for this to be true, it has to be the case that the "loss" is understood not in terms of preference satisfaction but in terms of the quantity of goods (in this case, primary social goods). As a result, this assumption is not really about the rationality of the parties but about the content of their desires. (See more about this under "Mutual Disinterest".)

g) Motivation: Desire for Primary Social Goods

- (1) By making the focus of the contractors be primary social goods instead of well-being, Rawls puts himself in the camp of resource theorists, as opposed to welfare theorists.

(2) Primary Social Goods

- (a) Definition: Rawls says that primary social goods are "things which it is supposed a rational man wants whatever else he wants" (p. 92, p. 70 in revised edition).

- (i) *First Bold Account*: Primary social goods are those goods that are directly or very largely under the influence of the basic structure and that are desired by every (rational) person as means to the fulfillment of his/her rational life plan. (See p. 93 in original edition:

“Now the assumption is that though men’s rational plans do have different final ends, they nevertheless all require for their execution certain primary goods, both natural and social. Plans differ since individual abilities, circumstances, and wants differ; rational plans are adjusted to these contingencies. But, whatever one’s system of ends, primary goods are necessary means. Greater intelligence, wealth and opportunity, for example, allow a person to achieve ends he could not rationally contemplate otherwise. The expectations of representative men are, then, to be defined by the index of primary social goods available to them. While the persons in the original position do not know their conception of the good, they do know, I assume, that they prefer more rather than less primary social goods. And this information is sufficient for them to know how to advance their interests in the initial situation.” (p. 93, original edition)

[a] *Problem:* Probably nothing fits the definition unless ‘rational’ is made to do more work than Rawls intends it to do.

[i] For example, beyond a very small amount, more wealth and income is not a benefit to the ascetic.

(ii) *Revised Definition (The Fall-back Plan):* Primary social goods are those goods that are directly or very largely under the influence of the basic structure and that are reasonably desired by the contractors because they are either a means to the fulfillment of every person’s rational life plan or at least having more rather than less of them is not harmful to such fulfillment. (See p. 142-3 in original, 123 in revised edition.)

[a] *Problem #1:* Probably nothing fits *this* definition unless Rawls presses ‘rational’ to do more than he intends.

[i] *Example:* The Weak-Willed Ascetic

[b] *Problem #2:* Furthermore, people certainly can be hurt by *others* having more rather than less liberty, for example.

(b) Extension:

(i) Rights and Liberties

(ii) Income and Wealth

(iii) Powers and Opportunities

(iv) The Social Basis of Self-Esteem

(3) Rawls’s Arguments for the Focus on Primary Social Goods Instead of Utility:

(a) *Argument from the Problems of Utilitarian Measurement (Or “Looking Where the Light is Good):* Utilitarianism requires cardinal and

interpersonally comparable measures of utility. There are problems defining and measuring utilities in such a way as to make this feasible and justifiable. Focusing on primary social goods simplifies the measurement problems.

(i) Comments:

[a] Rawls doesn't put much weight on this, and neither should we. Shifting to an object that is easier to measure but not of intrinsic moral significance is a questionable advance.

[b] Rawls's use of utilitarianism as a foil obscures the fact that the problems of measurement that a utilitarian encounters are partly the result of her acceptance of well-being as the currency for her theory and partly the result of her acceptance of a fully aggregative maximization theory.

(b) *Argument from the Veil of Ignorance*: Because the veil of ignorance has stripped the contractors of their real-life aims and ends, they cannot reasonably seek to promote their specific real-life ends and aims.

(i) Comments:

[a] This merely pushes back the problem of justification. Why should the contractors be denied knowledge of their real-life aims and ends? Rawls's general justification for the veil of ignorance restrictions is that the knowledge precluded is morally irrelevant and tends to undermine consensus. This strategy requires him to give a criterion of moral irrelevance. Furthermore, perfect knowledge of each person's aims and interests would not undermine consensus if each were precluded from knowing which person s/he is (at least given Rawls's other assumptions).

[b] It is a merely negative argument in that, if successful, shows why the contractors cannot directly promote their real-life aims and ends, not why they promote primary social goods.

(c) *Argument from Pluralism*: Rawls suggests that focusing on primary social goods instead of happiness is more compatible with the pluralistic commitments of liberalism (p. 94, revised edition pp. 80-1).

(i) *Comment*: This argument seems specious to me. Contractors would care about their *happiness* in real life—that is the reason they are motivated to pursue primary goods. The fact that they would want to set up institutions that did not engage in evaluating their real life ends and aims is irrelevant to the question of what criteria they are employing in choosing these institutions.

(d) *Argument from the "Doctrine of Necessary Means"*: This is probably the argument Rawls intended to put most weight on in the original edition. If he could get the extension of 'primary social goods' to come out as he intends using one of his definitions, he could argue that the contractors reasonably (in light of their ignorance) promote their shares of primary

social goods because they are in situation in which doing so is weakly dominant. That is, by doing so, they possibly gain something they care about but don't, in any event, risk losing anything they care about.

- (i) *Comment:* Unfortunately, the things he considers to be primary social goods don't fit his definition (even the "Plan B" definition). And probably nothing else does, either. If these things are not necessary means (or at least not ever impediments) to fulfilling one's life plans, Rawls argument is undermined.
- (4) Rawls's Arguments for Focusing on Primary *Social* Goods Instead of All Primary Goods.
- (a) *Argument from Feasibility:* Social goods are those that are either directly determined by the basic social structure (legal rights and liberties and legal powers and opportunities) or strongly influenced by that structure (income and wealth). Contractors are choosing principles for the basic social structure, so it is with respect to these goods that the contractors will judge that structure.
 - (i) *Comment:* As Rawls indicates, this is not a sharp or stable distinction. Many of the features of people that constitute primary natural goods may well be strongly influenced by the social structure (if not now, then in the future). If, for example, we could alter native intelligence *in utero*, should we do it for those with a bad genetic endowment but not for those with a good genetic endowment? Should people's body parts be subject to redistribution in accordance with the principles of justice?
- h) Mutual Indifference:
- (1) Rawls assumes that the contractors' preferences over outcomes are based only on what each can expect to get as individuals, not on what they get in relation to others. This is a very strong and controversial assumption.
 - (2) Rawls treats it as if this is an element of the contractors' rationality, but it really has to do with the *ends* they seek—a topic about which he said he wasn't going to make strong assumptions.
 - (3) This assumption simply begs the question against any form of true egalitarianism—*i.e.*, against any theory that says that the relative equality of people's resources or well-being, is of *intrinsic* moral significance. In effect, by making the assumption of mutual indifference, Rawls guarantees an "individualist" moral theory.
 - (4) He says he does it to keep the assumptions weak, but it is, in fact, a strong assumption. The assumption appears weak when we think of the project to be one of showing that even those with no benevolent motivation would agree to the principles in question. One might think that if rational, mutually disinterested individuals would agree to the difference principle, then *a fortiori*, more altruistically motivated people would, as well. But people aren't just motivated selfishly or altruistically. Some care about organic or holistic aspects

of a distribution. The assumption of mutual disinterest rules these people's concerns out of order in the reasoning for the principles of justice.

- (5) Furthermore, Rawls doesn't seem to appreciate that the primary goods are interrelated in such a way that concern for one's "absolute share" of some of them may require instrumental concern for one's *relative* share of others. For example, it is plausible to argue that one's absolute share of power depends on one's *relative* share of income and wealth and, further, that the absolute value of one's social bases of self-esteem depends, in part, on one's relative share of rights and liberties, powers and opportunities. If so, then in seeking the highest "index of [some] primary social goods" one must attend to one's relative share of other primary social goods.

3. Maximin Reasoning

a) Decision and Knowledge

- (1) *Decision Under Certainty*: One is in a situation of choice under certainty when one knows all of the consequences of each of one's alternatives. The received view of rational choice under certainty holds that it consists in maximizing utility. The standard economic conception of 'utility' takes it to refer to a measure of the agent's preferences.
- (2) *Decision Under Risk*: One is in a situation of choice under risk when one does not know the consequences of each of one's alternatives but one knows the *possible* outcomes of each alternative *and* the probabilities of each outcome conditional on each available alternative. Standard wisdom holds that rationality consists in maximizing expected utility, where the expected utility of an action is the probabilistically weighted sum of the utilities of each of its possible outcomes.
- (3) *Decision Under Ignorance (Uncertainty)*: One is in a situation of choice under ignorance when one does not know the consequences of each of one's alternatives, knows the *possible* outcomes of each alternative, but *does not know* the probabilities of each outcome conditional on each available alternative. There is no received view on what it is rational to do under such situations. It is quite possible that there is no correct *general* answer to the question of how it is rational to act in situations of ignorance. There are far more suggested strategies than Rawls considers, including:
- (a) *The dominance principle*: One action, a_1 , dominates another, a_2 , just in case in every possible state of the world the outcome of a_1 is at least as good as that of a_2 , and in at least one possible state of the world the outcome of a_1 is better than that of a_2 . The dominance principle holds that a dominant action should be preferred to one that is dominated by it.
- (i) *Problem*: Dominance reasoning is as near to being unimpeachable as anything in decision theory. However, the dominance principle only endorses a unique choice when there is one action that dominates all alternative actions and it is a rare situation when this occurs.

- (b) *The Principle of Indifference*: Maximizing expected utility using probabilities set by the principle of indifference (what Rawls calls the Principle of Insufficient Reason): The principle of indifference is a method for assigning probabilities when they are unknown. It holds that if there are n possible states of the world and you have no reason to believe any more likely than any other, you should assign a probability of $1/n$ to each possible outcome.
- (i) *Problem*: This method of fixing probabilities in ignorance produces the Bertrand Paradox. The paradox arises because there may be different methods of describing the possible outcomes that lead to inconsistent assignments of probabilities.
- [a] *The Bertrand Paradox*: Suppose that all we know is that a car finished a one mile track in between one and two minutes (inclusive). This means, of course, that its average speed was between 30 and 60 mph.
- [i] We have no more (or less) reason to think that it finished the track in between 1 and 1.5 minutes than we do that it finished in between 1.5 and 2 minutes. (The 1.5 minute mark divides the finish times into two equal parts.) Therefore, by the principle of indifference, the probability of the car finishing the track in between 1 and 1.5 minutes is $1/2$. And the probability of the car finishing the track in between 1.5 and 2 minutes is $1/2$. If it finished the track in exactly 1.5 minutes, its average speed was 40 mph. Therefore, the probability of the car's average speed being between 30 and 40 mph is $1/2$ and, therefore, the probability of the average speed being between 40 and 60 mph is $1/2$.
- [ii] We have no more (or less) reason to believe that the car averaged between 30 and 45 mph than we do to believe that it averaged between 45 and 60 mph. (The 45 mph mark divides the average speed range into two equal parts.) Therefore, by the principle of indifference, the probability of the car averaging between 30 and 45 mph is $1/2$ and the probability of it averaging between 45 and 60 mph is $1/2$.
- [iii] We have no more (or less) reason to believe that the car averaged between 40 and 45 mph than we do to believe that its averaged lay in any 5 mph range between 30 mph and 60 mph (*e.g.*, 30-35 mph, 35-40 mph, *etc.*). The average speed range can be partitioned into 6 such ranges so, by the principle of indifference, the probability of any of these is $1/6$. Therefore the probability of the average speed being between 40 and 45 mph is $1/6$.

In step [i], we calculated the probability of the car's average speed being between 30 and 40 mph as being $1/2$. In step [ii], we calculated the probability of the average speed being between 30 and

45 mph as $1/2$. This entails that the probability of the average speed being between 40 and 45 is 0. But in step [iii], we calculated this probability to be $1/6$.

- (c) *The maximin strategy (endorsed by Rawls for the specific situation of the original position)*: This strategy directs one to choose that action (or, in case of ties, any one of those actions) the worst possible outcome of which is at least as good as the worst possible outcome of every alternative action.
 - (i) *Problem*: This is an exceptionally cautious strategy—indeed, the most cautious conceivable. It is equivalent to choosing on the assumption that “nature is out to get you”—that no matter what action you perform, the world will be as bad for you as it could be. It focuses on minimizing possible losses (relative to alternatives) with no regard whatsoever for possible gains.
 - (d) *Insurance Strategies*: These strategies direct one to set a minimum satisfactory level and eliminate alternatives that risk falling below that level. Different insurance strategies offer different ways to choose between those actions that *do not* risk falling below the satisfactory level.
 - (i) *Problem*: There are, of course, problems with the rationale for setting the satisfactory minimum at any given level and the rationale for whatever principle is used to choose between the actions that do not risk falling below it. Furthermore, the satisfactory level could eliminate *all* available alternatives.
 - (e) *Maximax, Maxipenultimax, Maxipenultimin, etc.*: I offer these facetiously and leave it to you to figure out what they are and what the problems are with them. There *are* other serious proposals—the minimax regret principle, for example—but none have widespread acceptance.
- b) Rawls's Arguments for Maximin as the Rational Strategy for the Contractors in the Original Position
 - (1) *Argument from the Ignorance of Probabilities*: Rawls argues that the contractors lack the information needed to employ the expected utility principle. In order to use this principle, they would have to know the probability that they are represented by each of the representative men, presumably by knowing the size of the classes represented by each. The veil of ignorance precludes this knowledge. (See p. 154 in original, 134 in the revised edition)
 - (a) *Objection #1*: Rawls does not offer an adequate justification for construing the veil of ignorance broadly enough to preclude the knowledge in question. The number of people who are treated in a given way certainly seems to be morally relevant information. For a contrary view, though, see “Should the Numbers Count” by John Taurek (*Philosophy and Public Affairs* 6(1977) 293-316).
 - (b) *Objection #2*: Ignorance of the relevant probabilities places the contractors in a situation of decision under ignorance. But the maximin strategy is

only one of many strategies for choosing in such circumstances—and it is by no means the most plausible in all such situations.

(2) The Quasi-Dominance Argument:

(a) *Definition of Quasi-Dominance (N.B. this is not a standard decision-theoretic term):* One alternative, a_1 , is quasi-dominant with respect to another, a_2 , just in case in every state of the world the outcome of a_1 is almost as good as that of a_2 , and in at least one state of the world the outcome of a_1 is significantly better than that of a_2 . The quasi-dominance principle holds that a quasi-dominant action should be preferred to a quasi-dominated action provided that the probability of the state(s) in which the outcome of the quasi-dominant action is significantly better than that of the quasi-dominated action is *non-negligible*.

(b) Rawls's Assumptions:

(i) *Geometrically Diminishing Marginal Utility of Primary Social Goods:* Primary social goods are subject to geometrically diminishing marginal utility beyond a required minimum (p. 154; 134 in r.e.).

(ii) *Maximin's Unique Guarantee of the Required Minimum:* Use of the maximin strategy (and only the use of this strategy) will guarantee a satisfactory minimum share of primary social goods (p. 154; 134 in r.e.).

(c) *The Reasoning:* If the contractors' concern for receiving any payoff above the satisfactory minimum is very small compared to their concern for ensuring that they receive at least the minimum, and maximin, and only maximin, will guarantee not falling below that minimum, the contractors can use quasi-dominance reasoning for accepting the maximin strategy. In those states of the world in which other strategies would leave them below the satisfactory minimum (and there must be such states), the contractors have a great deal to gain by using maximin. In those states of the world where other strategies would offer the contractors some gain over what the maximin strategy would yield, the gain is insignificant to the contractors. Assuming that the probability of winding up below the satisfactory minimum by using other strategies is non-negligible, the quasi-dominance principle would direct the contractors to employ the maximin strategy.

(d) Justification of the Assumptions:

(i) *Geometrically Diminishing Marginal Utility of Primary Social Goods:* Rawls doesn't offer anything that could really be considered a justification for this claim. It seems ludicrously strong.

(ii) *Maximin's Unique Guarantee of the Required Minimum:*

[a] *Maximin's Guarantee:* Rawls is never clear about this initially puzzling claim. It is not as if we can change the states of the world by selecting a different strategy for choosing between the alternatives we have.

- [i] *The Circumstances of Justice:* It is here in Rawls's argument that the assumption of the circumstances of justice plays a crucial role. The assumption of moderate scarcity ensures that there is some way for goods to be distributed so that each has a satisfactory minimum. If so, then use of the maximin strategy would ensure that everyone receives a satisfactory minimum.
 - [b] *The Uniqueness of Maximin's Guarantee:* Rawls doesn't even discuss this, though it is obvious that if there is some distributive schema that accords to each a satisfactory minimum, then there are many strategies that could guarantee that such a distributive scheme would be selected. For example, an insurance strategy would do so. The difference is, of course, that if there are several distributive schemas that would guarantee to each a satisfactory minimum, an insurance strategy could select a different one than that selected by the maximin strategy.
- (e) *The Importance of the Argument:* The quasi-dominance argument is clearly the most interesting argument Rawls offers for the maximin strategy. There is reason for thinking that he doesn't think it is essential.
- (i) *The Lexical Difference Principle:* Rawls accepts what he calls a *lexical* version of the difference principle (p. 83, 72 r.e.). This principle involves the iterative use of the difference principle to break ties that may exist after a single employment of it. It works as follows: if there are two (or more) arrangements that maximize expected life-prospects of the worst off representative person, then we are to choose that arrangement that maximizes the life-prospects for the next worst off; if there is still more than one alternative remaining, select the one that maximizes the life-prospects for the *next* worst off, *etc.*.
 - (ii) *The Lexical Maximin Strategy:* The lexical difference principle would be selected by the contractors only if they choose in accordance with a lexical maximin strategy.
 - (iii) *The Effect on the Role of the Quasi-Dominance Argument:* The lexical maximin strategy cannot be justified by the quasi-dominance argument because, for any level above the satisfactory minimum, Rawls cannot consistently argue (as he must) that the contractors' concern for any gain above that level is very small compared for their concern to ensure that they not fall below that level.
- (3) *Argument from Acceptability of Institutions:* Rawls argues that alternatives to the maximin strategy would lead to institutions that are unacceptable.
- (a) *Reflective Equilibrium Interpretation:* The plausibility of this as a reflective equilibrium argument depends, of course, on what sort of institutions one finds acceptable in reflective equilibrium. Many have found institutions that do not accord with the difference principle to be acceptable. Even people who want a secure "social safety net" think that

the net need not be set as high as possible (as the difference principle seems to imply it must).

(b) *Stability Interpretation:* Rawls might be arguing that from the point of view of the original position, the institutions that would be selected by following another strategy would not be acceptable because they would be unstable.

(i) *Criticism:* This argument is specious. To the degree that the contractors care about the stability of their institutions, this is already reflected in the “primary goods matrix”. The contractors care only to secure for themselves the highest share of primary social goods. Therefore, if stable institutions (or institutions that are acceptable in real life) are important, this importance is already reflected in the primary goods numbers.

4. *From the General Conception of Justice to the Two Principles (and their Ordering):* The maximin strategy leads directly to the general conception of justice (final statement: p. 303; p. 266 revised edition). But it is less clear how Rawls gets from the general conception to the famous two principles of justice—especially the first.

a) *The General Conception of Justice:* “All primary social goods ... are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored.”

b) *The Two Principles:*

(1) “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.”

(2) “Social and economic inequalities are to be arranged so that they are both:

(a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and

(b) attached to offices and positions open to all under conditions of fair equality of opportunity.” (Statement from p. 302, p. 266, revised edition.)

c) *The Priority Rules:* I won't retype them here, but the first priority rule ranks the first principle of justice “lexically” prior to the second. This means that it is never just for a society to be organized in such a way as to not maximize liberty (compatible with like liberty for all) no matter what gain in other primary goods it can achieve for the least well off. It also means that it is never just for a society to accept inequality in liberty no matter what gain in other primary social goods *or liberty, itself*, can be achieved for the least well off.

d) *Attempt to Justify the Derivation of the Two Principles from the General Conception:*

(1) *The Zero-Sum Game Argument for the First Principle of Justice:*

(a) *Two Theses:* Here are two theses which, if both true, would help to justify the derivation of the first principle of justice from the general conception of justice:

- (i) *Contiguity of Rights Thesis*: This thesis holds that (it is always true that) one person's rights and liberties end where someone else's begins. Rights always "touch" one another and leave no unclaimed space.
 - (ii) *Fixity of Moral Space Thesis*: This thesis holds that the amount of space for rights and liberties is "fixed". As real estate investors say when they are trying to explain why real estate is a good investment, "God ain't making any more!"
- (b) *The Argument*: The above theses guarantee that the choice of various schemas regarding the distribution of rights and liberties is always a "zero-sum game." This means that one person's gain is always another's loss. If this is true, there can be no "efficiencies of inequality" and, in particular, the worst off can never benefit from inequality. Therefore, the general conception of justice leads directly to the first principle.
- (i) *Criticism*: As it is, this is inadequate for the conclusion. Even if the decision concerning the distribution of rights and liberties is a zero-sum game, inequality of rights and liberties may increase the total share of *all* primary social goods enjoyed by the worst off.
 - [a] *Reply*: The priority of liberty (the first priority rule) precludes making these trade-offs.
 - [i] *Note*: This reply shows that the argument for the first principle of justice *depends on* the argument for the priority rule. That is, one must argue, first, that rights and liberties are special and are ranked lexically prior to the other primary social goods before one can argue that rights and liberties must be distributed equally.
- e) *The Argument for the Second Principle of Justice*: Unlike the "derivation" of the first principle, it is clear enough how the second principle follows from the general conception of justice with two possible "hitches."
- (1) The Just Savings Principle; and,
 - (2) The requirement that positions be open to all.
- f) *The Argument for the Lexical Priority of the First Principle*: Appeal to intuitive plausibility of the conclusions reached by assuming the principle.
- (1) *Criticism*: People's intuitions differ here. Rawls has been criticized for parochialism and an attitude toward the relative value of the primary goods that seems plausible only to those who are relatively privileged. (See below for more on the first priority rule.)
5. Problems with the Difference Principle
- a) Applicability Problems
 - (1) Utilitarianism's Problem of Interpersonal Comparisons

- (a) Rawls says he isn't going to stress these problems with utilitarianism, but he believes that the Difference Principle ameliorates the problems.
- (2) *Rawls's Boast*: The Difference Principle encounters less serious problems with the problem of interpersonal comparisons for two reasons.
 - (a) What Rawls says here is very confusing. Here is his claim: "...[A]s long as we can identify the least advantaged representative man, only ordinal judgments of well-being are required from then on" (p. 91, 79 r.e.). But this seems to run together two separate issues: the distinction between cardinal and ordinal measures and the difference between interpersonal and intrapersonal comparisons. Identifying the least advantaged representative man does not require a cardinal measure of well-being. It does, though, require an *interpersonal* ordinal measure. After identifying such a person, we can measure his improvement and disimprovement in terms of purely *intrapersonal* ordinal measures. However, if Rawls thinks that, after such identification, we can be done with *interpersonal* ordinal evaluations, he is wrong. For every social system we consider—that is, for every principle we consider—we need to find the worst-off representative individual. To do that, we must compare the well-being of each individual in society to that of every other individual. Then, we must compare the well-being of the worst-off representative individual in each social scheme to the worst-off individual in each other scheme. That is yet another *interpersonal* comparison—though still an ordinal one.
 - (i) *A Reconstruction*: In applying the Difference Principle, we needn't make cardinal judgments of well-being (or levels of primary social goods) at all. We need only ordinal judgments though they will be *interpersonal* ordinal judgments.
 - [a] This doesn't go all the way in solving the problem.
 - (b) The basis of comparisons is expectations of primary social goods.
 - (i) This is what solves the problem. But notice: this isn't the result of moving from a theory that requires one to maximize total or average good to one that requires one to maximize the good of the worst off. It is the result of moving from a theory of 'good' in terms of subjective states of the agent to one that understands 'good' in terms of things that are objectively measurable. If a utilitarian adopted such a theory of the good, (for example, if she adopted a hedonistic theory or one in terms of income) she wouldn't have any problem with interpersonal comparisons of individual good, either.
 - (ii) Employing primary social goods as the currency for evaluating the justice of social institutions raises its own problem of comparisons: not interpersonal comparisons but the problem of comparing different goods. This is the problem faced by all pluralistic theories: how and in terms of what currency, does one compare various goods? Rawls invokes the priority of liberty assumption and an assumption about the cohesion of primary goods to "solve" this problem.

- [a] The priority of liberty “solution” has (as Bertrand Russell says in another context) “all the attractions of theft over honest toil.”
 - [b] The “cohesion of primary goods” thesis, if true, seems to be a case of Rawls having his theory “saved by a lucky fact”.
- b) *Plausibility Problems:* (Many of the criticisms of Rawls's Difference Principle will surface as we examine Nozick's and Gauthier's criticisms. Here is one that won't come up in those discussions.)
- (1) *Scarcity:* I have argued that the scope of justice is broader than Rawls believes. In particular, I have argued that there are just and unjust ways to distribute goods that are extremely scarce relative to demand. If so, then an adequate theory of justice must give answers about how to do this—answers that are acceptable in reflective equilibrium.
 - (a) *The Unacceptability of the Difference Principle in Cases of Extreme Scarcity:* In cases of extreme scarcity, adherence to the difference principle will result in (as one critic has said, “equality in nothing—or something so close to nothing that no one could care about it”).
 - (b) *Rethinking Scarcity from a Rawlsian Perspective:* As I hinted before when talking about the scope of justice, I believe that reflecting on extreme scarcity, while it undermines the universality of Rawls's two principles, can actually support Rawls's hypothetical contractarian methodology. This is because I'm inclined to think that contractors in a Rawlsian-like original position would choose principles to handle extreme scarcity that would be acceptable in reflective equilibrium.
 - (i) *A Suggested Principle for Extreme Scarcity:* If we believe that contractors would be very risk averse (even if not as absolutely risk averse as Rawls believes), there is reason to think that they would agree to principles of distribution of extremely scarce, essential resources that would maximize (or nearly maximize) the number of individuals who wind above a satisfactory minimum. This is a “not implausible” principle for dealing with extreme scarcity.