

Joseph D. Sneed

Political Institutions as Means to Economic Justice: A Critique of Rawls' Contractarianism

Abstract: It is argued that John Rawls' theory of social justice as well as the contract argument for it are misleading, if not actually mistaken, in that they appear to take institutional features of societies as fundamental objects of moral evaluation. An alternative view is expounded. Principles involving institutional features are only contingently related to principles involving the distribution of things people care about. These distributions are taken as the fundamental objects of moral evaluation. Social, political and economic institutions are means to achieve more desirable distributions. It is argued that the alternative provides a more accurate reconstruction of the moral foundations of social-democratic liberalism than does Rawls' theory.

Introduction

In *A Theory of Justice* (Rawls 1971) John Rawls has offered us an account of what societies *ought* to be like. In this account two aspects of societies are relevant to their moral evaluation: (a) certain features of a society's basic institutional or organizational structure; (b) the way things that people care about are distributed among members of a society. These aspects of a society are "lexicographically" ordered in the sense that societies are first rank-ordered according to a criterion involving (a), and then societies that are equivalent in this rank-ordering are rank-ordered according to a criterion involving (b). Rawls argues for the ethical plausibility of this evaluation procedure by an iterated application of refined version of traditional "social contract" theory. Roughly, a contract argument is first applied to derive a principle of distributive justice that rank-orders societies according to (b)-features. Then, under the assumption that all "contracting parties" accept this principle of distributive justice, a contract argument is employed once again to derive a principle that rank-orders societies according to (a)-features AND the lexicographical priority of the second principle with respect to the first.

My basic aim in this paper is to demonstrate that this theory of social justice as well as the contract argument for it are at best misleading, if not actually "mistaken", in that they appear to elevate institutional features of societies to fundamental objects of moral evaluation. I believe the "correct" view is that principles involving institutional features are only contingently related to principles involving the distribution of things people care about. These distributions are the fundamental objects of moral evaluation — social, political and economic institutions are means toward achieving more desirable distributions. By "correct" here, I mean that one obtains a more coherent, convincing and faithful reconstruction of the moral foundations of social-democratic liberalism when this distinction is clearly drawn than when it is obscured or ignored, as I find it to be in Rawls' theory¹.

My criticism focuses on Rawls' iteration of contract arguments. First, I attempt to lay out the general logical structure of a contract argument with particular attention to separating and identifying its moral and non-moral premises. Second, I indicate what I find to be serious, but perhaps not devastating, difficulties with Rawls' first application of this argument form to derive a principle of distributive justice. Next, I argue that my interpretation of Rawls' argument for his full theory of justice as a two-stage iteration of a contract argument is accurate. This interpretation appears to be plausible only under the assumption of a contingent connection between "just" social institutions and "just" distributions. But, under this assumption, the moral character of the second contract argument vanishes and its plausibility appears dubious. Further, the second argument is seen to rest on some rather strong assumptions about people's preferences. Somewhat different, equally strong, but no less plausible, assumptions about preferences appear to sufficiently derive the full Rawls theory from his principle of distributive justice without appeal to a contract argument. This suggests that the contingent connection between Rawls' full theory and his principle of distributive justice is quite weak and that use of a contract argument to make this connection obscures this weakness. Examples are mentioned which illustrate the weakness of this connection.

The General Principle of Distributive Justice

To understand my criticism we need to have a rough understanding of the content of Rawls' general principle of distributive justice. Rawls states this principle in slightly different ways in two places:

All social values — liberty and opportunity, income and wealth, and the bases of self-respect — are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage. (Rawls 1971, 62)

All social primary goods — liberty and opportunity, income and wealth, and the bases of self-respect — are to be distributed equally unless an unequal distribution or any of all of these goods is to the advantage of the least favored. (Rawls 1971, 303)

By "social values" and "social primary goods" Rawls means those "things" that people care about which they receive by virtue of their position in the class structure of society. Roughly, these are their social position *per se*, supposing people do in fact care about this, and the expectation of receiving other things they might care about which is associated with their class.

This principle invites us to envision a variety of ways in which the same group of people with access to the same resources can be assigned to mutually exclusive social classes. In every configuration of social classes, each class has some role in the economic activity of the group — does some kind of work, in the most general sense — and consumes some share of product of this activity. The intuitive content of the principle is roughly this. Allowing some members of the society to work at intrinsically more satisfying jobs or consume more in "compensation" for their work

may increase the society's total output of "good things" over what it would be if everybody were deriving the same satisfaction from their work and consumption. If this increase in total output is distributed so that it raises the level of everybody's satisfaction — not just the "more productive" — then the departure from equal distribution of satisfaction is justified.

It is important to understand that the general principle of justice is a principle for evaluating the "output" of social institutions — the distributions of things people care about — not the institutions themselves. Institutions are evaluated directly only to the extent that people may have intrinsic preferences for certain institutional forms or intrinsic preferences for certain positions in institutions apart from the expectations of getting other things they care about that are associated with this position. The principle rank-orders institutions only indirectly by looking at the distributions they produce. In contrast, Rawls' full theory focuses in part at least on pointing out desirable features of institutions themselves. One may regard the full theory as telling us something about the kinds of institutions that are likely to produce distributions consistent with the general principle of justice. Rawls is not clear about this distinction. In a sense I am "foisting it upon him". His main concern is with saying something about institutions likely to produce just distributions — with the special principles of justice. This is surely an intelligible enterprise. But it can be distinguished from the question of what just distributions are. Further, I think it ought to be. Mainly because I am considerably more confident about how to formulate my own views on what distributions are just than I am that I shall ever be able to say anything very strong about what kinds of institutions are most likely to produce these distributions.

Let us now turn to Rawls' argument for the general principle of justice. It is convenient to consider first the general form of the argument and then respectively the underlying theory of rational decision making and the characterization of the decision problems that belong to it.

The Contract Principle

Rawls believes that understanding the subjective conditions under which people would unanimously prefer living in a society organized in accordance with a given principle of justice (meaning that people actually behave in the way the principle dictates) to living in societies organized in accordance with other principles sheds light on the content of the given principle. These conditions he calls "the initial state" or "the initial situation". He believes there is a correspondence between "initial situations" and preferred principles of justice:

We may conjecture that for each traditional conception of justice there exists an interpretation of the initial situation in which its principles are the preferred solution . . . The procedure of contract theories provides, then, a general analytic method for the comparative study of conceptions of justice. One tries to set out the different conditions embodied in the contractual situation in which their principles would be chosen. In this way one formulates the various underlying assumptions on which these conceptions seem to depend. (Rawls 1971, 121-122)

Not only are we enlightened about the content of the principle corresponding when we know its initial state, but in some sense we come to see the “underlying assumptions” on which the principle depends. This suggests that Rawls believes the corresponding initial situation is somehow more enlightening about the intuitive plausibility of a particular principle than are its implications in specific situations.

At least part of what the corresponding initial situation tells us about a principle of justice has bearing on the ethical plausibility of the principle. Rawls thinks it is possible to identify situations where it is intuitively plausible to think that people would choose morally “correct” principles of justice:

I assume . . . that there is a broad measure of agreement that the principles of justice should be chosen under certain conditions. To satisfy a particular description of the initial situation one shows that it incorporates these commonly shared presumptions. (Rawls 1971, 18)

The idea is apparently this. Not only do we have ethical intuitions about what principles of justice entail for concrete applications; we also have ethical intuitions about the corresponding initial situations. Thus one way to argue for the ethical plausibility of a principle of justice is to demonstrate that it corresponds to an ethically plausible initial situation.

This way of arguing for the ethical plausibility of principles of distributive justice I call “the contract principle”. This is not Rawls’ term. Indeed, it is somewhat difficult to distill from his extensive discussion of “the contractarian argument” (mainly in Rawls 1971, Ch. III) a precise statement of the logical form of this argument. But I believe that my notion of the contract principle does capture what he has in mind. The essential idea of the contract principle is a correspondence between individual decision situations – situations in which an individual with a specific configuration of beliefs and preferences is confronted with a choice of living in societies organized in accordance with alternative principles of justice – and the principles of distributive justice that would be chosen in these situations. The possibility of using this correspondence to construct an argument for principles of justice hinges on whether people do in fact have strong ethical intuitions about the circumstances in which morally correct principles of justice would be chosen. Whether an argument constructed in this way is more or less compelling than an argument based on tracing out the implications of the principle in specific examples depends upon whether people have firmer intuitions about the circumstances in which morally correct principles of justice would be chosen or firmer intuitions about what is just in specific situations. Of course, by happy coincidence both these lines of argument might lead to the same principle. Rawls clearly thinks this to be the case for his general principle.

It is important to understand that the contract principle is a *moral* principle. It translates a *moral* rank-ordering of initial situations (via non-moral principles of rational decision theory) into a *moral* rank-ordering of principles of distributive justice. It does not permit one to derive a *moral* rank-ordering from a descriptive one. One may always ask, “Why *should* I rank-order the initial situations in the way Rawls suggests”.

It is possible to criticize arguments based on the contract principle at two levels. One can attack these arguments in general, maintaining perhaps that *no* argument of this form could be compelling. Or one can attack specific examples of such arguments like Rawls' contractarian argument for the general principle. There appear to be two major lines of criticism that could be pursued against contractarian arguments in general:

1. No suitable correspondence between individual decision problems — initial situations — and principles of justice like that required by the contract principle can be established;
2. People do not in fact have coherent ethical intuitions about which initial situations will result in the choice of ethically plausible principles of justice.

Successful pursuit of any one of these lines of argument would appear to vitiate contractarian arguments in general.

I shall not attempt to carry out any of these lines in detail, but only indicate briefly what would be involved. To see what line (1) involves, think of the contract principle in the following way. We have given a rank-ordering of initial situations according to their ethical plausibility. We have some general account of individual rationality — a theory of individual decision making. This theory maps initial situations into principles of justice. Rawls apparently thinks that this mapping will be one-one onto the set of "traditional conceptions of justice". If it is, there is no problem with ordering the set of conceptions of justice in the same way as their corresponding initial situations. Even if the mapping is one-many, we can still order the set of principles of justice. The only difference is that now we may not have a single unique principle of justice corresponding to the top-ranked initial situation. But if the mapping is many-one — several initial situations correspond to the same principle of justice — then we will not in general be able to order the set of principles of justice in the same way as their corresponding initial situations. Suppose initial situations A, B and C are rank ordered as

A
B
C,

A and C both correspond to principle P, while B corresponds to principle Q. How do we order P and Q?

It seems pretty clear that in general the mapping from initial situations into principles of justice is going to be many-one. "Trivial" modifications in a given initial situation would correspond to the same principle of justice. If these modifications — trivial from the point of view of a theory of individual rationality — were "non-trivial" from the point of view of ethical plausibility — if they were ranked differently — then the correspondence required by the contract principle would not exist. This construction provides a general method of attacking contractarian arguments. Suppose A (above) is the top-ranking initial situation from the point of view of ethical plausibility and B is ranked just below it. Attempt to construct some C which corresponds to the same principle of justice as A but is clearly ethically inferior even to B. I shall suggest below that a construction of this type may well vitiate Rawls' contractarian argument for the difference principle.

It is clear that in many initial situations there is no reason to expect that people would choose morally correct principles of justice. In most situations that could conceivably actually occur in which people were confronted with choices among alternative principles of justice we might expect them to choose principles which they believe are likely to favor them. The contractarian argument invites us to abstract from these situations and consider situations in which lack of information prevents people from effectively identifying and pursuing their own interests in choosing principles of justice. The basic idea is that the conception of justice Rawls is trying to capture is one which would command unanimous assent if only people could abstract from their own self-interest. But, a critic pursuing line (2) might argue, the concept of "abstracting from their own self-interest" is a bit vague. There are a number of ways one can characterize initial situations which appear to correspond to this — a number of ways of characterizing absence of information needed to identify and pursue one's own interest. Each of these may lead to a different principle of justice. And while all are clearly more intuitively likely to lead to a choice of a morally correct principle of justice than situations that might actually occur, our intuitions fail when we try to distinguish among the abstract initial situations.

In a weaker version of line (2), the critic would admit that we can intuitively rank-order the rather abstract initial situations where people are substantially ignorant of their own interests. But he would maintain that our intuitions here are rather feeble. The whole idea of highly contrived decision situations which lead to different principles of justice depending upon precisely how you interpret "abstracting from their own self-interest" is a bit fanciful. While it may be possible to carry through an argument in this manner, the conclusion is somewhat suspect. In particular, were it to lead to a principle of justice that conflicted with our intuitions about what is just in a concrete situation, we would give up the principle.

The Maximin Principle

The main idea of Rawls' contractarian argument for the difference principle is to describe some initial situation which has two properties:

1. The plausibility of people in this situation choosing the correct principle of justice is greater than for any other situation.
2. It would be reasonable for people in this situation to employ the maximin principle of individual decision making.

The maximin principle states roughly: consider the worst that can happen to you in each alternative open to you; choose an alternative for which the worst is as good or better than the worst for any other alternative. From property (2) it is trivial to argue that people would choose Rawls' general principle in this initial situation. It is important to note that this only suffices to establish the general principle of justice (Rawls 1971, 151). Further argument is required to establish the specific principles of justice.

The first thing to note here is that whether the maximin principle is an independent principle of rational decision making is an issue in some dispute. Many Bayesians would argue that all situations in which the maximin principle is plausible can be handled with the expected utility principle by attributing relatively high probabilities and low utilities to the "worst" alternatives. That is, using the maximin principle is justified if and only if the situation warrants a certain kind of systematic pessimism which can be described in terms of subjective probability and utility values. Rawls apparently thinks there are some situations where the maximin principle applies in which its applicability cannot be deduced from a "more fundamental" expected utility principle:

Clearly the maximin rule is not, in general, a suitable guide for choices made under uncertainty. But it is attractive in situations marked by certain special features . . . First, since the rule takes no account of the likelihoods . . . , there must be some reason for sharply discounting estimates of these probabilities . . . for example, that the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure . . . second . . . the person . . . cares very little, if anything, for what he might gain above the minimum stipend . . . third . . . that the rejected alternatives have outcomes that one can hardly accept. (Rawls 1971, 153-154)

Rawls' last two conditions are easily translatable into conditions on the utility function. But his first condition indicates that he thinks there is a class of situations in which the maximin principle should be used in preference to the expected utility principle and our "best estimate" of the subjective probabilities. Thoroughgoing Bayesians would deny this.

From the thoroughgoing Bayesian's point of view, if our estimates of probabilities are so imprecise as Rawls describes, then the best rational decision theory can do is weed out the obviously irrational alternatives. It cannot yield a unique solution to the decision problem².

Depending upon the ranges one is ready to assume for the probability values the maximin solution may or may not be included in the set the theory does not dismiss as clearly irrational. Only if one makes rather strong "pessimistic" assumptions will the maximin solution be uniquely selected. In general though, there will be a wide class of not clearly irrational principles of distributive justice — one corresponding to each plausible probability distribution attributable to persons in the initial state. Indeed the "maximize average utility" principle will be in this set so long as the uniform equal probability distribution is among those that it is plausible to think persons in the initial state might have.

Rawls is concerned in particular to argue that in the initial situation he describes the "maximize average utility" principle would not be chosen while his general principle would (Sec. 27). Further, he wants to argue that in his initial situation it is more plausible that morally correct principles of distributive justice would be chosen than in situations in which the "maximize average utility" principle would be chosen. Now, if you are a thoroughgoing Bayesian whether or not the "maximize expected utility" principle will be chosen depends upon whether or not the uniform equal probability distribution is among those it is plausible to think of people

in the initial situation as having. If you are not a thoroughgoing Bayesian then it depends upon whether the initial situation is one of those decision problems where the expected utility principle does not apply. In either case it is rather difficult to see how the way these questions are resolved can effect the intuitive plausibility of the situation's being one in which morally correct principles of distributive justice would be chosen.

It is difficult for me to see how the question of whether or not the inequities sanctioned by the "maximize average utility" principle are morally justified can intuitively depend upon these rather delicate questions of what the range of people's subjective probability distributions are and what is the best theoretical account of how they should use this information to make decisions. Suppose we accept the claim that morally correct principles of justice would be chosen in those situations in which people do not have enough information to identify and act on their own self-interest. It is still hard to see how this rough description of an initial situation can be plausibly translated into the conceptual framework of rational decision theory in a way that will yield solutions to the decision problem unique enough to be interesting. What the theory yields depends on which version you choose to use — Bayesian or non-Bayesian — and precisely how you render the intuitive notion into the theory — what kinds of probability distributions or lack thereof you choose to represent the intuitive notion. I remain unconvinced that the way of doing this which yields the difference principle is the obviously correct one. Roughly speaking, the output of rational decision theory here in terms of principles of justice is highly sensitive to the input in terms of subjective probability and utility functions. It is surely possible, as Rawls demonstrates, to find some way to put these intuitions into the language of the theory that yields our favorite principle. But by tinkering slightly with essentially technical details we can get the theory to yield a variety of other principles including one we find morally unacceptable. This suggests to me that rational decision theory is much too sensitive an instrument for the rather crude ethical intuitions we have in this context. I doubt that one could ever plausibly employ the contract principle to yield essentially new results — principles we were not confident about in advance.

The Initial State

Rawls' characterization of "the favored philosophical interpretation" of the initial situation — the one in which it is most plausible to think that people will choose the correct principles of justice may be divided into three parts:

- a characterization of the alternatives (Rawls 1971, Sec. 21 and 23)
- a characterization of the information available (Sec. 24)
- a characterization of other morally relevant features (Sec. 22).

We have already considered the "information situation". This is what allegedly justifies the application of the maximin principle.

The characterization of alternatives is a justification for confining the set of principles to those that have some historical significance (Sec. 21) and imposing some formal constraints like universality, generality, simplicity, publicity, non-dictatorship and transitivity of their ordering of distributions of primary goods (Sec. 23). Some of these like transitivity and generality seem to be simply a part of what we mean by a principle of justice. Other so-called formal constraints like non-dictatorship have substantive content. Given the strength Rawls appears to find in his characterization of the information state — he thinks it justifies using the maximin principle — it is hard to see why he needs these substantive constraints on the alternatives. A maximiner would presumably never opt for a dictatorial principle.

The problem here seems to be that Rawls is not clearly distinguishing here between initial situations in which people are being asked to choose from among several alternative candidates for a general principle of justice and initial situations in which they are being asked to choose among several alternative systems of rules of behavior or systems of basic social institutions designed to implement a previously agreed upon general principle of distributive justice. He makes this distinction clearly when he comes to talk about the realization of the special principles of justice in a constitutional democracy (Sec. 31). He ignores it here apparently because he wants to provide one single contractarian argument for both the general principle and the special principles. But he does not in fact do this. The argument of Ch. III has as its conclusion *only* the general principle of justice. This is made explicit in the last paragraph of p. 151. The argument (Sec. 82) for the priority of liberty, though contractarian in form, must be understood in quite a different way.

More important for the argument for the general principle is his characterization of the other morally relevant features of the initial state — “the circumstances of justice”. Rawls divides these into “objective circumstances” — such things as the geographical proximity of the participants in the initial situation, comparable mental and physical powers, moderate scarcity of natural resources — and “subjective circumstances” — roughly that the preferences of the participants are selfish but such that some kind of mutually profitable cooperation is possible. Formally what Rawls has described is a multilateral non-zero-sum game with a possibility for introducing cooperation.

It is not clear to me how essential the non-zero-sum property is to the claim that people in the initial situation Rawls is describing would choose the correct principle of justice. However, it is pretty clear that the non-zero-sum property could be dropped without rendering the claim of the applicability of the maximin principle less plausible. If it is reasonable to maximinimize in the information situation Rawls describes in the context of a non-zero-sum game, then it is clearly reasonable to maximinimize in the same information situation in the context of a zero-sum game. Now suppose that it is in fact less plausible to think that people would choose morally correct principles of justice in a zero-sum situation. Indeed, suppose that it is more plausible to think that they would choose morally correct principles in a non-zero-sum situation with enough information to assign a uniform equal probability distributions to the positions in the class structure than it is to think that they would

choose correct principles in a zero-sum situation with an information situation that warrants maximinimizing. Were this the case then a suitable correspondence between initial situations and principles of justice would not exist. We would have a situation like that described above (p. 129) in which A and C are respectively the Rawls initial situation and this situation with the non-zero-sum requirement dropped while B is the Rawls situation modified by assuming just enough information to assign uniform equal probabilities. A and C lead to the Rawls principle, while B leads to the maximize average utility principle. We would have no basis for rank-ordering the difference principle and the maximize average utility principle.

This is not offered as a “knock down” objection to the contract argument for the Rawls principle. I can see no very persuasive reasons to think that people would be less likely to choose correct principles of justice in a zero-sum, maximinimizing situation than in a non-zero-sum equal probability situation. But I admit to having rather weak intuitions about the circumstances in which people would be likely to choose principles of justice that appeal to me on other grounds. Indeed, I think *this* is the trouble with the contract argument. When you see the form of the argument laid out clearly, it requires us to rank-order initial situations as to the “likelihood” that people in these situations would choose morally correct principles of justice. The conclusions of the argument are very sensitive to the way we order not very different initial situations. Unless one has rather sharp views about the fine details of situations in which correct principles would be chosen, the contract argument is not likely to be conclusive.

Let us now turn to Rawls’ full theory of just societies.

The Specific Principles of Justice

Rawls begins to trace out the implications of his general principle of justice by noting the form it would take in a wide class of specific situations. The derived or “specific” principles of justice are:

First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others.

Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage, and (b) attached to positions and offices open to all. (Rawls 1971, 60)

These are reformulated in a slightly different way on p. 250 and p. 301 which makes more explicit Rawls’ view that the first principle is prior to the second. These principles presuppose a distinction between political and economic institutions in a society:

... these principles presuppose that the social structure can be divided into two more or less distinct parts ... They distinguish between those aspects of the social system that define and secure the equal liberties of citizenship and those that specify and establish social and economic inequalities. (61)

By “basic liberties” Rawls means:

... political liberty (the right to vote and to be eligible for public office) together with freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person along with the right to hold (personal) property; and freedom from arbitrary arrest and seizure as defined by the concept of the rule of law. (61)

The first principle requires that such liberties be extended equally to all citizens, while the second principle:

... applies, in the first approximation, to the distribution of income and wealth and to the design of organizations that make use of differences in authority and responsibility, or chains of command. (61)

The first principle is prior to the second in the sense that:

... a departure from the institutions of equal liberty required by the first principle cannot be justified by, or compensated for, by greater social and economic advantages. The distribution of wealth and income and the hierarchies of authority, must be consistent with both the liberties of equal citizenship and equality of opportunity. (61)

Political liberty takes precedence over economic justice.

On my interpretation of Rawls’ argument, the “theory of society” plays no direct role in deriving the special principles from the general principle of justice — only assumptions about individual preferences seem to be invoked. The theory of society seems to be employed to describe social conditions in which individuals are likely to have preferences of this sort. There is some difficulty here in separating the factual from the normative claims. The most plausible interpretation I can construct is this. The theory of society is normative³.

In societies satisfying this normative theory individuals will (as a matter of fact) have preferences of the sort needed to derive the special principles from the general.

That these special principles of justice are indeed to be regarded as following from the general principle under certain further assumptions is clearly stated.

It is clear that these principles are rather specific in their content, and their acceptance rests on certain assumptions that I must eventually try to explain and justify. A theory of justice depends upon a theory of society in ways that will become evident as we proceed. For the present, it should be observed that the two principles ... are a special case of a more general conception of justice ... (62)

What is less clear is the epistemological status of these “further assumptions”.

The important thing to note about the specific principles of justice is that the first specific principle is about the properties that social institutions should have which are apparently unrelated to their outputs. The second principle — essentially the general principle restated — is still about the output of social institutions. The force of the special principles is to tell us that in certain circumstances certain

non-output-related features of social institutions are *always* going to be a part of any institutional structure that produces just distributions of things people care about. The “always” here could be taken in two ways:

- as indicating necessary connection — just distributions are in these circumstances restricted by definition to those produced by institutions with these properties;
- as indicating contingent connection — in these circumstances just distributions are in fact only produced by institutions with these properties, or less strongly, institutions with these properties are more likely than any others to produce just distributions.

I have opted for the second interpretation, though I am not entirely certain that this is the one Rawls has in mind.

The source of my uncertainty is this. Rawls sometimes talks as if there were some contract argument for the special principles that bypassed the general principle of justice. Almost the whole of Chapter III with the exception of the paragraph beginning at the bottom of p. 151 appears amenable to this interpretation. However, I believe his argument for the special principles in Sec. 82 supports my view that he regards this as a contingent connection. The argument is contractarian in form and it appears clear that Rawls regards the plausibility of the initial situation to hinge on assuming that those choosing institutional forms have come to a prior agreement on the general principle of justice:

Having chosen a conception of justice that seeks to eliminate the significance of relative economic and social advantages as supports for men's self-confidence, it is essential that the priority of liberty be firmly maintained. So for this reason too the parties are led to adopt a serial ordering of the two principles. (545)

I take this to mean that the parties are envisioned to be choosing kinds of institutional forms intended to *implement* a previously agreed upon principle about just distributions.

Let us now consider, in detail, Rawls' contract argument for the special principles of justice.

The Contract Argument for the Special Principles of Justice

Rawls' argument for the priority of equal distribution of political liberties over economic justice based on his general principle of distributive justice (Sec. 82), has the same structure as his contract argument for the general principle, but the alternatives being considered are different kinds of things. The general principle is a principle for rank-ordering the distributions of things that people care about independently of the institutions that produce them — except that people may have intrinsic preferences about their roles in institutions. In the contract argument for this principle people in the initial situation are faced with a choice among the Rawls general principle and other principles which similarly rank-order the “outputs” of social institutions. In contrast to this sort of decision problem, the contract argu-

ment for the priority of liberty invites us to consider a decision problem in which people in the initial situation are faced with a choice of the *kind* of basic institutions that will be present in the society. "Kind" is emphasized here to distinguish this choice from the choice of a particular political constitution. What I am suggesting is that "the first stage" of Rawls' "four stage sequence" of contractarian arguments (Sec. 31) be broken into two parts: one in which general principles of justice are rank-ordered; and a second in which general features of institutions designed to implement the highest ranked principle will be rank-ordered.

The same general objections to contract arguments raised earlier still apply here. But there is an additional one which seems to bear in this case. What is at issue in the present case is basically an empirical question. It is a question of political science, not political doctrine. Which sorts of social institutions are most likely to realize a given principle of distributive justice is not a "matter of moral opinion" in a sense that the question of what is the correct principle of distributive justice is. Looking at what people might do in hypothetical situations in which they lack information enough to rationally pursue their own interests *may* shed some light on moral judgments about how the things people care about ought to be distributed. But it is significantly less plausible to think that doing this would be a very fruitful way to pursue empirical social science — even very "theoretical" empirical science. The contract argument here is somewhat analogous to considering what a group of physicists who all accepted Newton's laws might do if they got together to "choose" a force law to account for the motion of the planets in ignorance of any specific empirical data about the paths of the planets.

This seems to me to be a fundamental difficulty with Rawls' derivation of the special principles of justice. The contract principle is only plausible as a normative principle. Once one has seen clearly that the argument of the special principles involves a two-stage iteration of the contract argument it is possible to ask of each iteration whether the contract principle is plausible. In the first iteration where the conclusion is a moral judgement the principle may have some plausibility. In the second iteration where the conclusion deals with means of implementing a moral judgment the contract principle smacks of "arm-chair social science"⁴.

But, even granting the plausibility of the contract principle here, quite "heroic" assumptions are needed to push the argument through.

To understand Rawls' argument here the notion of liberty must be made a little more precise. Roughly, liberty is a property of a system of prescriptive rules about people's behavior (see Kanger 1972, 125 ff.).

A person P has liberty with respect to decisions of kind K in the system of rules S if and only if the rules of S entail nothing about P's behavior when confronted with decisions of kind K.

When we speak of "the priority of equal liberty" over economic justice we are saying something about "optimal" systems of rules. At the very least, we are saying roughly that some system of rules in which everybody has liberty with respect to some kinds of decisions is at least as "good" as any other set of rules. Good for

what? Presumably good for implementing the preferred general principle of distributive justice. Thus a weak version of the priority of equal liberty claim would be this.

— There exist a kind of decisions K and a system of rules S such that:

1. All people have liberty with respect to K in S;
2. with respect to decisions not of kind K the rules of S are consistent with the difference principle;
3. general compliance with S is at least as likely to produce distributions consistent with the general principle of justice as general compliance with any other system of rules S'—i.e., S is optimal with respect to the general principle.

Stronger versions of this claim would involve specifying more or less precisely the kind of decisions K with respect to which all people have liberty in S. Rawls clearly argues for one of these stronger versions in which K is those areas protected by the basic civil liberties of a constitutional democracy.

It is also clear that Rawls believes that not just anything can be a system of rules. The discussion in Rawls 1971, Sec. 23 is most appropriately regarded as describing a set of "feasible" rules from among which the people in the initial situation must choose. The preceding statement of the weak priority of equal liberty claim can be easily modified to require that all rules considered be in the feasible set.

It is clear that there is at least one "system of rules" that is sure to satisfy (3) in the priority of equal liberty claim. The rule is simply: in all situations of conflict among persons people shall do as the difference principle prescribes. Here people have liberty only with respect to decisions which other people do not care about. Presumably this system of rules though "optimal" is not feasible. But it is worth contrasting with a system of rules in which some people have liberty with respect to some kinds of decisions other people care about. This contrast suggests that there are really two questions at issue here. First, is there *any* "optimal" system of rules in which people have liberty with respect to socially non-trivial kinds of decisions — decisions that other people care about? And second, if so, are there any of these systems in which *all* people have liberties with respect to some of the *same* socially non-trivial kinds of decisions? Rawls wants to establish that in all feasible optimal systems of rules some people will have liberty with respect to some socially non-trivial kinds of decisions. The only question for him is whether any of these liberties are possessed by everyone in a feasible system that comes as close to producing difference principle justice as any other feasible system.

For the moment let us confine our attention to the weak priority of equal liberty claim and ignore the restriction to feasible rules. One may distinguish at least two ways to argue for this claim:

1. Accept the Rawls principle as characterizing distributive justice and argue that in certain circumstances this principle entails the priority of equal liberty claim;
2. Argue that in certain initial situations where people who accept the difference principle possess slightly more information about their preferences than in those situations required to establish the difference principle, people would choose a system of rules having properties (1) and (2) in the priority of equal liberty

claim. And further, it is plausible to think that systems chosen in this situation would satisfy (3).

Rawls' argument is clearly of the form of (2). But it is instructive to compare it to a somewhat parallel argument of form (1).

The key to both these ways of arguing for the priority of equal liberty is characterizing:

- in (1) the circumstances in which the difference principle entails the priority of equal liberty;
- in (2) the initial situations in which people would choose a system of rules with the requisite properties.

Rawls clearly envisions himself to be arguing in way (2). In doing so he characterizes certain situations in terms of general propositions about the kinds of things people want. He clearly intends this to be taken as characterizations of an initial situation. But they might plausibly be taken as characterizations of circumstances in which the difference principle entails the priority of equal liberty.

To see this let us look a bit more closely at the assumptions needed to make the second iteration of the contract argument sound.

General Propositions about the Kinds of Things People Want

Rawls appears to assume that people in the initial situation in fact have rather strong preferences for exercising discretion – making their own decisions – in his terminology “determining our own plan of life” (Rawls 1971, 543). He mentions two reasons why this is so. First, people have “... moral, religious or philosophical interests which they cannot put in jeopardy ... which they must keep themselves free to honor” (206). That is, people want to be in a position to act in accord with their moral, religious and philosophical insights – as perceived by them. A second, “... is the central place of the primary good of self-respect and the desire of human beings to express their nature in a social union with others” (543). Roughly, what Rawls seems to mean here is that people have a strong preference for social situations which reinforce their own conception of their importance to other people and perhaps even in “the grand scheme of things” – situations in which people have or are thought by other people to have the prerogative of discretion on certain matters of this sort.

Rawls does not think that these intrinsic preferences of liberty dominate all others. The satisfaction of certain basic needs is preferred to liberty, but

... as the general level of well-being rises (as indicated by the index of primary goods the less favored can expect) only the less urgent wants remain to be satisfied by further advances, at least insofar as men's wants not largely created by institutions and social forms. At the same time the obstacles to the exercise of equal liberties decline and a growing insistence upon the right to pursue our spiritual and cultural interests asserts itself. (542-543)

That is, once the basic needs are met then liberty is preferred to the satisfaction of all less basic desires.

There appears to be some ambiguity in Rawls' discussion here. Is he claiming that people in the initial situation believe that when the "veil of ignorance" is lifted they will have strong *specific* preferences in some general areas? Or is he claiming that in the initial situation they have strong preferences for general discretion in these same areas? His reference to "moral, religious and philosophical interest" appears to support the first interpretation while his reference to "the primary good of self-respect" appears to support the second. Either of these interpretations will suffice for the contractarian argument (2). Only the second, applied to actual situations, rather than the hypothetical initial situation will suffice for the direct argument (1).

In an actual situation in which all people care about discretion in making decisions of kind K more than anything specific they might do in *any* non-K situation, the Rawls general principle clearly entails the priority of equal liberty claim. Restricting the liberties afforded one social class could not be justified on the grounds that it raised the level of the lowest social class. For this class would regard their loss of liberty as placing them in a worse position than they were initially no matter how many more of their non-K preferences were satisfied in the new situation. Given these assumptions, this is a straightforward consequence of the general principle. No contract notions are required to establish it. It is to be noted, however, that the assumptions about preferences needed to establish it are quite strong. Indeed, given these assumptions the conclusion is hardly surprising. If people in fact strongly prefer K-liberty to other things, then any plausible principle of justice, even mere Pareto-efficiency, will demand they be given K-liberty. A rather egalitarian principle of justice unsurprisingly recommends an equal distribution of K-liberty in such circumstances.

Given that this rather straightforward argument from the general principle appears to be available to establish the priority of equal liberty, why should one resort to a contract argument? One might claim that in actual choice situations people do not in fact have strong preferences for discretion in general areas. Rather they have strong preferences for specific things in this area. The reason for moving from concrete situations to an ideal initial situation is to abstract from people's specific strong preferences and consider only their beliefs about the *kinds* of things they care strongly about.

For example, I (and a partner) may be consumed by a passion to copulate in the city hall fountain at high noon on the 4th of July, but not necessarily by a passion for general discretion in matters of sexual conduct. We can still apply the general principle to determine what is just here. I should be permitted to copulate in the city hall fountain if and only if preventing me makes me worse-off than those whose sensibilities would be offended were I permitted to do so. One might think this to be a rather weak case for my being permitted to do what I want, depending as it does on mere contingencies about relative strengths of preferences. Indeed, is it really just to allow me to copulate in the city hall fountain and forbid you to do so

simply because you will not enjoy it as much as I? Would not a contract argument which justified equal distribution of sexual liberty be preferred?

Certainly people in the initial situation, being ignorant of their specific preferences about sexual conduct, might still agree that this was an area of activity where their preferences were likely to be quite strong – whatever their specific nature. To prevent the thwarting of these preferences they might agree to reserve discretion in this area for individuals. They might even agree to this even though it turned out later that the widespread discomfort occasioned by my copulating in the city hall fountain far outweighed the moderate satisfaction it afforded me. The Rawls principle applied in a specific conflict situation might give a different answer than a rule derived from the maximin decision rule applied in the initial situation to a choice among systems of rules for resolving this sort of conflict.

People in the initial situation *might* choose a system of rules providing for equal absolute liberty of sexual conduct. But would they? Surely behind “the veil of ignorance” they must also countenance the possibility that *their* sensibilities may be strongly offended by the choices that some people will make in the area of sexual conduct. To determine whether those in the initial situation would opt for equal absolute liberty of sexual conduct we must know how they rank the worst that can happen to them with this liberty – witnessing the most reprehensible conceivable “perversion” – and the worst that could happen to them under each feasible system of rules that entailed some restrictions on sexual conduct – the thwarting of the strongest conceivable desire whose satisfaction might be forbidden by that system of rules.

The issue here of interest to us is not whether they would choose equal absolute liberty of sexual conduct or some more restrictive rule. The issue is: would whatever rule they choose extend the same liberty of sexual conduct to all? It is tempting to think that the following principle could be invoked to prove that they would choose a rule that extended the same liberty of sexual conduct to all.

In a situation of ignorance as to whether or not they possess property P there is always some conflict resolution rule that does not discriminate on the basis of P which a maximinimizer would prefer to any given rule that does discriminate on the basis of P.

Suppose there are two P-non-discriminatory rules R_1 and R_2 with m_1 and m_2 respectively the worst conceivable outcome under each rule, and m_1 preferred by everybody to m_2 . Construct the discriminatory rule: “ R_1 applies to people with property P; R_2 applies to people with non-P.” Clearly maximinizers would find the discriminatory rule no better than R_2 and clearly inferior to R_1 if they were completely uninformed about whether or not they were P or non-P. This is apparently the principle that lies behind Rawls’ construction of the initial situation. He has an intuitive idea of the kinds of properties that are irrelevant to the distribution of liberties and he makes the people in his initial situation ignorant of these things. Given this principle, it follows that people in Rawls’ initial situation would not choose a principle of sexual liberty that discriminated on the basis of properties they did not know whether they had in the initial situation.

But is this principle generally true? Suppose R_1 permits general copulating in the city hall fountain while R_2 generally forbids it. Suppose m_1 is having a strong aversion to witnessing copulation in the city hall fountain and being forced by circumstances to do so, while m_2 is being forbidden to satisfy an overpowering desire to copulate in the city hall fountain. Suppose P is the property of having an overpowering desire to copulate in the city hall fountain. The discriminatory rule now permits those who *really* want to copulate in the city hall fountain to do so and forbids others. A maximinimizer would clearly prefer it to R_2 and would find R_1 no better. Indeed, departing from the maximin rule, one might argue that by restricting copulating in the city hall fountain to those who really want to do it we diminish the probability that those who have aversion to witnessing such activities would actually be exposed to them. Thus the discriminatory rule might be strictly preferred to both R_1 and R_2 . It is not too difficult to construct a discriminatory rule that even a maximinizer would strictly prefer to both R_1 and R_2 . Simply require those permitted to do things like copulate in the city hall fountain to pay some compensation to those who have an aversion to witnessing such things.

The point is this. The principle which seems to lie in back of the argument that people in the initial situation would choose non-discriminatory systems of rules appears not to work for rules that discriminate on the basis of preferences people might have. One could argue at this point that rules which discriminate on the basis of preference are not feasible. But this appears to be generally false. Why not regulate the privilege of copulating in the city hall fountain by the sale of permits? One might even envision more sophisticated schemes in which the proceeds from the sale of permits were employed to compensate those offended by the spectacle (see, for example, Massel 1971).

Of course this does not prove that there are not some areas of activity which people in the initial situation would choose to regulate by rules that make no discrimination on the basis of information not available in the initial situation. What assumptions about an area of activity would be sufficient to assure this? Not simply that people in the initial situation believed that they would have strong preferences about specific decisions of this kind. This held in our example about sexual conduct and was clearly insufficient. Suppose we added to this the assumption that they did not believe they would be very much damaged by anything other people might do in exercising discretion in this area. This is clearly sufficient and even justifies equal *absolute* liberty in the area. But it is also clearly too strong to be of much interest. Areas of activity where I care very much about what *I* do may unfortunately also be areas in which I care very much about what *you* do.

We might try supposing that people care enough about decisions in this area that they cannot be compensated for thwarted preferences. But this is insufficient too if they care about what other people do and can be compensated for putting up with their doing it—i.e., compensation and their doing it is strictly preferred to no compensation and their doing it. More strongly we might suppose that no compensation either way is possible. Your offence at my satisfying my desire can no more be assuaged by compensation than my frustration at its being thwarted. This clearly

rules out discriminatory principles involving compensation. But it is quite strong. In the area of Rawls' favorite example — religious liberty — it is hard to believe that I could not find *some* way to compensate you for the offence my religious practices gave you, even though perhaps you could not compensate me for suppressing these practices. Even if we rule out discriminatory rules involving compensation with this assumption, the argument for ruling out other discriminatory rules is still rather weak. It depends on the fact that the maximin rule ranks a discriminatory and a non-discriminatory rule together while a Bayesian rule could plausibly rank the discriminatory rule higher. Unless we are willing to hinge the whole argument for the priority of liberty on *this* application of the maximin rule, it appears that we must fall back on the assumption that people simply have strong preferences for discretion in certain areas. This suffices for the contract argument. But it also suffices for the non-contract argument directly from the general principle.

This difficulty appears again when Rawls comes to deal with specific examples of political liberties.

He begins tracing out the implications of the special principles of justice by considering equal distribution of basic political liberties. He looks at three examples: equal liberty of conscience (Rawls 1971, Secs. 33, 34 and 35); equal right to participate in the political processes (Secs. 36 and 37); and equal treatment under judicial and administrative systems (Sec. 38).

The discussion of equal liberty of conscience closely parallels the general discussion of the priority of liberty in Sec. 82. The same ambiguity between people's having strong preferences for general discretion and people's believing they will have strong preferences in "matters of conscience" pervades it. Some evidence that Rawls favors the "strong preference for general discretion" line is provided by his discussion of what a utilitarian argument for equal liberty of conscience might be like:

... a utilitarian may try to argue from the general facts of social life that when properly carried out the computation of advantages never justifies such limitations . . . But even if the parties were persuaded of this, they might as well guarantee their freedom straightaway by adopting the principle of equal liberty. There is nothing gained by not doing so, and to the extent that the outcome of an actuarial calculation is unclear a great deal may be lost. (207)

Rawls is apparently speaking here of an argument based on the average utility principle of distributive justice and assumptions about strong preferences for general discretion. It is not obvious why an argument based on his general principle of distributive justice and similar assumptions about preferences would not be strong enough to satisfy Rawls. The only thing the contractarian argument seems to "add" is foreclosing the possibility revising public policy on the basis of new factual information about preferences. Why it would be rational for people in the initial situation to "lock themselves into" a policy of equal liberty of conscience — or anything else — is simply not clear.

Along with this contractarian argument for equal liberty of conscience a kind of "inductive" argument can be extracted. It runs like this. Our intuition about

what is just indicates that equal liberty of conscience is just and that it takes precedence over issues of economic justice. This is consistent with the special principles of justice — in particular, the priority of the principle of equal liberty. Therefore we have a “confirming instance” of the specific principles of justice.

One might readily agree with Rawls’ intuitions about the priority of equal liberty of conscience, yet question its persuasiveness as a confirming instance of specific principles of justice. Liberty of conscience, construed narrowly as the right to hold and publicly proclaim religious, moral and philosophical views is a rather favorably chosen example. Restricting this kind of liberty for some social classes, one might argue, is not likely to have much effect on what the economy produces. The possibility of the worst-off being made economically better-off by such a restriction is simply not very likely to arise. Indeed, the justification for equal liberty of conscience might be largely independent of how strong people’s preferences for liberty of conscience happen to be. People could in fact care very little about such things and the Rawls principle would still demand equal liberty of conscience because nothing was to be gained for the worst-off by doing otherwise. On the other hand, if you choose a somewhat richer conception of liberty of conscience — say freedom from “indoctrination” by the major institutions of the society — our intuitions may not be so unequivocal. The systematic undermining of tribal mores in developing countries is sometimes justified on economic grounds. This could conceivably be consistent with the Rawls principle — say, if the urban workers were actually worse-off than rural tribesmen who were the subjects of indoctrination. It is not obvious to me that a good social-democrat would find such indoctrination is unjust.

In discussing the merits of equally distributed political liberties Rawls again ultimately bases his argument on the alleged fact that people have strong intrinsic preferences for such liberties. He stipulates that:

The principle of equal liberty, when applied to the political procedure defined by the constitution, I shall refer to as the principle of (equal) participation. It requires that all citizens are to have an equal right to take part in, and to determine the outcome of, the constitutional process that establishes the laws with which they are to comply. (Rawls 1971, 221)

and goes on to say equal participation means representation on the basis of “one man, one vote” (223) and that its extent is determined by the extent to which simple majority rule is decisive in legislation (224). Rawls then considers (Sec. 37) arguments for limiting the extent of equal participation. Constitutional limitations on simple majority rule he agrees may be justified by the need to protect “more basic” liberties like “liberty of conscience and freedom of the person”. But equal participation is not to be regarded as simply a means to protect these liberties and even less as a means to secure economic justice. He rejects the suggestion that social classes, rather than individual persons might be the appropriate entities to be represented in the legislature as well as Mill’s suggestion that representation should be apportioned on the basis of intelligence and education. He argues:

... The grounds for self-government are not solely instrumental ... the effect of self-government where equal political rights have their fair value is to enhance the self-esteem and the sense of political competence of the average citizen ... Since he is expected to vote, he is expected to have political opinions. The time and thought that he devotes to forming his views is not governed by the likely material return of his political influence. Rather it is an activity enjoyable in itself that leads to a larger conception of society and to the development of his intellectual and moral faculties. (233-234)

No argument is presented to support the claim that participation in the political process is intrinsically valuable to everyone, nor even that it would be in a society well-ordered by the specific principles of justice.

More serious here than failure to support sweeping claims about what people want is the failure to deal with real issues of choice between equal political participation and economic justice. For example, civil servants and military personnel in some countries are barred from full participation in the political process. Might it not be reasonable to do the same for key executives – or even all employees – in corporations doing more than 75 percent of their business with the government? Is the distribution of representation in the U.S. Senate on a geographic basis really unjust supposing that it in fact results in some redistribution of the benefits of public spending from areas of high per-capita income to areas of low per-capita income. Might not the redistributive impact of taxation be improved (judged by the Rawls principle) if votes were distributed in inverse proportion to pre-tax income? Is representation based on socio-economic classes really unjust? Might not some representation scheme on this basis conceivably be the most efficient way to maintain a distribution of income among these classes ranked highest among the politically feasible distributions by the Rawls principle? Rawls' discussion here is remarkably deficient in failing to deal seriously with intuitively plausible counter examples to his view. Instead, he emphasizes an argument based on very strong, highly questionable and essentially unsupported claims about people's preferences for political participation.

Notes

- 1 I have argued elsewhere (Sneed, to appear) that Rawls' theory may be understood as an attempt to provide a "rational reconstruction" of the moral foundations of this political movement, rather than an attempt to provide a universally valid criterion for moral evaluation of societies. I do not know whether Rawls would accept this interpretation. It could be that my limited interpretation of the theory's range of application is essential to the plausibility of my criticism. It could be that just the features I find objectionable are essential to bringing the views of some non-social-democrats within the theory's scope. But, I would argue that including these features excludes significant aspects of the social-democratic tradition so that one obtains neither a universal theory, nor one encompassing the social-democratic tradition.
- 2 For a discussion of how this is done see Raiffa 1968, Ch. 6.
- 3 There is also a descriptive theory of society implicit in Rawls' entire approach to social justice I have dealt at length with this point in Sneed 1976 but I do not think this is what he has in mind here.

- 4 The same objection can be made with even more force for the further three iterations of Rawls' "four-stage sequence" (Rawls 1971, Sec. 31). At each iteration "parties to the contract" are faced with deciding how best to pursue goals agreed upon in the previous step. These are always factual questions and in many cases quite technical (e.g., what tax policy is most effective in implementing a given principle of distributive justice). That the decisions which *would* result from an idealized "referendum" on such matters could shed *any* light on such policy decisions seems unlikely.

Bibliography

- Kanger, S. (1972), Law and Logic, in: *Theoria* 32, 105–132
 Massel, B.F. (1971), Externalities, Public Goods, Allocative Efficiency and Compensation: A New Approach. Discussion Paper No. 7, Institute for Public Policy Analysis, Stanford University
 Raiffa, H. (1968), *Decision Analysis*. Reading, Mass.
 Rawls, J. (1971), *A Theory of Justice*. Cambridge, Mass.
 Sneed, J. (1977), John Rawls and the Liberal Theory of Society, in: *Erkenntnis* 10, 1–19
 — Some Consequences of Rawlsian Justice in Simple Economies, to appear