Is Pareto Optimality a Criterion of Justice?

A *pareto optimal* social state is a distribution of goods such that the goods could not be reallocated to make anyone better off without making at least some person worse off. A situation in which each person prefers her bundle of goods to any other possible bundle is pareto optimal, but so is the situation in which I have all of the goods (and they are all of some, even the most minimal, value to me) and no one else has anything. Thus, considered alone, pareto optimality does not guarantee justice in the distribution of goods on anyone's account of justice. However, a pareto non-optimal distribution seems to cry out for improvement, since a reallocation can make at least one person better off without making anyone else worse off. Indeed, the individuals who would be advantaged by these improvements might claim that it is only just to require them. So pareto optimality also does not seem unconnected to justice. In this paper I examine arguments for the claim that justice requires pareto optimality.[1]

To be clear, I am asking whether pareto optimality is a *necessary condition* of justice in the distribution of goods. The situation in which any one individual has all the goods, which is pareto optimal provided that each good has some positive utility for that individual, shows that pareto optimality cannot be a sufficient condition for justice. But it remains an open question whether an ideally just society would have to satisfy pareto optimality, or whether pareto optimality is at least a regulative ideal for actual societies seeking justice. The set of pareto optimal distributions normally has many members, and thus is too indeterminate to give us an adequate theory of justice by itself on formal grounds alone. So we have to combine it with other criteria simply to get a full theory of distributive justice. Not all possible criteria can be combined with pareto optimality, however. If pareto optimality is

incompatible with other necessary criteria of justice, then we shall have to reject pareto optimality.

Support for the claim that pareto optimality is a criterion of justice has come mainly from three directions: utilitarianism, welfare economics, and contractarian moral theory. On the utilitarian view, a society is just if it maximizes the total or average utility of the society, and on either account this implies that a just society must meet the condition of pareto optimality. Though maximization of utility generates a pareto optimal distribution, utilitarians do not directly argue for it; pareto optimality is simply a by-product of maximization of utility. For this reason, as well as because I think utilitarianism does not provide an adequate account of justice, I shall not consider utilitarianism further. Welfare economics treats pareto optimality as a matter of individual rationality, a matter of individuals preferring more utility to less. Social choice theories commonly assert pareto optimality as a condition of social rationality as well. Contractarianism, especially in the theory of John Rawls, holds an ambivalent relationship to the pareto optimality condition. Rawls claims that justice is prior to pareto optimality ("efficiency" in his terms), yet he seems to require it in arguing for the lexical version of the difference principle. Another contractarian, David Gauthier, presents an unambiguous argument for pareto optimality as a condition of impartiality in the perfectly competitive market, and then goes on simply to assume that it ought to be a condition of justice outside that model.

I shall maintain that all of the arguments that have been put forward for pareto optimality as a condition of justice rely on some crucial idealization that makes them uninteresting arguments for the actual world. In making this claim I am not dismissing the usefulness of idealized models or hypothetical states of nature. Both kinds of structures can teach us much about our world. But not just any model or hypothetical state will do. The problem I mean to raise is sometimes termed by economists "robustness": a model is not robust when small changes in some parameter (assumption) cause very large changes in the results. A model that is not robust cannot be a good approximation to reality. If a model lacks robustness, then it is not very useful unless it accurately

represents, in that aspect of the model that lacks robustness, the reality one is trying to capture. Thus I shall argue that the existing models used to argue for pareto optimality as a criterion of justice are not robust.

I shall also argue that pareto optimality nevertheless provides a regulative ideal for a just society. My argument is a contractarian argument, since I take it that justice arises as a set of agreed upon constraints on individuals' self-interested behavior. Only contractarianism aims to show that the concerns of justice are rational for all to adopt and adhere to. The argument I shall offer does not depend on these claims about contractarianism, although the practical value of the argument may. In adopting the constraints of justice, I show that rational individuals will choose to regulate themselves by the ideal of pareto optimality, and this will persuade them to look to a wide range of alternative possible futures in making social policies. Constraints on what individuals can know about each other and their preferences and about possible futures will lead me to argue that a contractarian theory of justice demands democratic processes and dialogue to assure agreement and compliance. The pareto criterion, as it can be applied in the real world, will turn out to encourage progressive social policies.

The paper contains three sections. The first section examines arguments for the pareto criterion from the perfectly competitive market (PCM) model, including the welfare economists' rationality argument and Gauthier's impartiality argument. I shall show there that the arguments for the pareto criterion are not robust to realistic adjustments of the idealized model from which the argument is made. The second section examines arguments for the pareto criterion from Rawls's original position (OP). I argue that the hypothetical initial situation from which the arguments are made also relies on crucial, false assumptions that cannot be relaxed without compromising the conclusion that pareto optimality is a criterion of justice. In the third and final section I attempt to make the strongest argument for the pareto criterion that avoids these idealizations. Critical examination of this argument shows that it also fails to be robust. However, arguments made on the way lead me to conclude that nonetheless pareto optimality is a progressive regulative ideal in a just society.

1. The Argument for Pareto Optimality from the PCM

The concept of pareto optimality, or efficiency as it is sometimes called,[2] was invented by Vilfredo Pareto,[3] an Italian economist at the turn of the century, as a way to compare states of affairs without making the interpersonal comparisons of utility that utilitarianism requires. Welfare economists avoid interpersonal comparisons of utility mainly because they claim that utility is essentially private. Thus they avoid making claims about total utilities in society, fearing that they are meaningless sums. The pareto criterion has three advantages over maximization of total or average utility for a theory of justice. First, one can determine which state(s) are pareto optimal without making interpersonal comparisons of utility. All that matters for judgments about pareto optimality are individuals' own preferences for different social states. Second, pareto optimality is a weaker criterion than maximization of total or average utility, and unlike them does not generally provide a unique ideal distribution. Since many different distributions may satisfy pareto optimality, holding pareto optimality to be a criterion of justice need not imply that one holds a welfarist[4] conception of justice; pareto optimality can be a necessary criterion of justice in a view that takes other, non-welfarist criteria to be equally necessary. Furthermore, believing that ethical judgments are beyond their expertise, the only bases left by which welfare economists could make comparative social policy judgments are pareto optimality and related criteria.[5]

Welfare economics has made remarkable progress with these weak criteria. The two most important results of welfare economics concern an idealized model of a market called the perfectly competitive market (PCM). PCMs are characterized by the following conditions:

C1. Resources are privately owned and privately consumed.
C2. Force and fraud does not exist.
C3. There are many buyers and sellers, and entry and exit to the market is free, so that no one can act unilaterally, or collusively, to affect prices.

C4. Transactions (e.g., publicizing or gathering information about goods for exchange, shipping, preparation for consumption) are costless.
C5. Individuals' utility functions are stable, monotonic (which means that if any amount of a good ever raises an individual's utility, no greater amount of the good would lower the utility), and reflect transitive preference orderings.
C6. There are no externalities, that is, all costs and benefits of producing or consuming a good are borne by the owner of the good.
C7. No transactions take place out of equilibrium.[6]

Of course, the PCM model abstracts from reality in each of these conditions. The condition that goods are privately owned and consumed insures that no one takes another's interest as her own; each acts in her own self-interest. This condition implies what is sometimes called "non-tuism," and it rules out not only altruism, but perhaps more importantly envy, as motivations for individual actions. The requirement that utility functions be monotonic is pretty accurate for most goods, especially when there is a resale market. But it is well documented that individuals' preferences often fail to be transitive.[7] There are externalities in almost every economic activity, and in the world these give rise to free riding, public goods, political posturing, and lawsuits. Ideally a society might be able to internalize many of them, but there will always be some externalities. C7 requires that everyone know what equilibrium prices are, and that any price below the equilibrium price would be rejected by the seller, any price above by the buyer.

The other three conditions of the PCM rule out strategic behavior, or acting in ways to bluff or confuse the other individuals in the market. By prohibiting force and fraud from playing a role, the PCM rules out taking the market to be a larger game, in which illicit, extra-market behavior is figured in as a possible strategy, one with great risks and great potential rewards, to be considered in the light of rational self-interest. If transactions were costly, then there might be an advantage in delaying or hiding information so that some advantageous trades would not be made. Finally, the condition that no one can affect prices rules out monopolies and oligopolies, so that there is no incentive for individuals to

cooperate in small coalitions; in the PCM it is each person for herself.

In the PCM, individuals trade until all trades that are mutually agreeable to the buyer and the seller have been made, since trades are costless, they can get something they want better by trading, and they have no fraud to worry about. The first fundamental theorem of welfare economics tells us that in the PCM, individuals seeking their own interests will reach a pareto optimal outcome.

WE1: In the PCM the outcome of free trade is pareto optimal.

This can be shown informally by supposing that the outcome is not pareto optimal. Then there is at least one person who can be made better off without making another worse off. Suppose that person is Ed and he could be made better off by trading two of his oranges to Evelyn for one of her apples, a trade that would not make her worse off. Then (provided that there are no externalities) by making the trade they make a *pareto improvement,* that is, they increase someone's utility without decreasing anyone else's, and neither will see any reason not to.[8] If there are any other potential pareto improvements then they are made as well, by the same kind of argument. If not, then there is no way to make anyone better off without making another worse off, that is, the situation is pareto optimal. Pareto optimal states are those in which the distribution of goods is as effectively used to raise total welfare as it can be, and in the PCM it is done without forcibly taking anything from anyone. There are many such outcomes that we would not call just, in particular if the starting point for trade was unfair or coercive. Still, there is a connection to justice for anyone who thinks that liberty[9] is one component of it: WE1 implies that pareto optimality is a byproduct of the PCM, in which force and fraud as well as government coercion is absent.

The second fundamental theorem of welfare economics extends the significance of pareto optimality for concerns of justice.

WE2: Any pareto optimal outcome can be achieved through the PCM from some initial allocation of goods.

In other words, if one chooses a particular pareto optimal allocation, say for reasons of equity, it could be reached through free trade from some particular initial allocation, which can be arrived at through lump-sum taxes and forced transfers. Beginning from a fair initial allocation of goods, the PCM allows agents to come to a new allocation, which everyone likes even better, through freely entered trades, and this maximizes the social welfare that can be achieved given the starting point. Since WE2 allows that any pareto optimal end-state allocation could have been brought about through the liberty of the free market, welfare economists take it that pareto optimality is at least a weak criterion of justice, justifying the market as the only just allocative mechanism. As Hal Varian writes, "the interesting result of welfare economics is that we can relate an end-state principle of justice—maximum 'social welfare'—to an allocative procedure—the market mechanism."[10] This fact has often been used, controversially, to justify the free market in the real world.

The view of most welfare economists is that since pareto optimality can be had without sacrificing equity or other distributional concerns, then it would be foolish not to take it. Who could complain about a pareto improvement? This seems to be a powerful appeal (though as stated not an argument from justice). But it faces several key objections. First, the argument for pareto optimality as a criterion of justice from WE2 does not extend beyond the confines of the PCM. In the real world the conditions of the PCM do not exist, and neither WE1 nor WE2 hold. Perhaps most significant is that there exist externalities. The existence of externalities means that there will be beneficial trades that are not made, and costs imposed on persons who are not party to trades. The classic examples are lighthouses and pollution. Lighthouses provide a valuable service to ship navigators, for which they would be willing to pay if that were the only way that they could use it. But the beam of light that the lighthouse sends is free for all who see its beam to consume, so naturally, shippers will try to consume it without paying for it. Knowing this, few would want to construct lighthouses in a private free market, since they will bring about financial loss for most of them.[11] Hence there will be an undersupply of lighthouses in the free market. In just the opposite

way, too much pollution will be produced. Firms who pollute a river, for example, are able to let their waste, which would otherwise be costly to dispose of, run downstream freely, where it becomes someone else's cost. The firm then can sell its product without considering that cost, and the consumers of the product will consume it without having to pay that cost. Since some of the consumers would not pay the higher cost which includes waste disposal, more of the product, and hence more of the pollution, is created than there would be if the waste disposal cost were internalized. The result of externalities is that the free market does not lead to pareto optimality. Provided that there are enough shippers who would pay something for it, if they could prevent any non-payer from consuming it, a lighthouse could be constructed which would make all of them better off without making anyone else worse off. And provided that the pollution is costly enough to the downstream neighbors, if they could prevent the firm from freely using the river, a level of pollution that reflects the costs and benefits to all could be chosen. But since there are externalities in the world, and some of them cannot be internalized, the link between the freedom of the PCM and pareto optimality is broken.[12] What is more, there is no proof that pareto optimality can always be had without sacrificing equity, since WE2 does not hold in the presence of externalities.

The welfare economists' argument tries to play both sides of the libertarian-egalitarian debate. It claims that the PCM achieves pareto optimality, and cites the PCM's lack of force and fraud in appealing to the libertarian. It cites WE2 to show that an egalitarian pareto optimal solution can be achieved through the market in appealing to the egalitarian. But without some coercion, usually in the form of lump-sum taxes and transfers, there is no guarantee that the outcome of the untouched PCM will be just by the egalitarian's standards (or those of anyone but the libertarian). So either one sacrifices freedom (in the sense of the libertarian) for equality, or equality (in the sense of the egalitarian) for freedom.[13]

The welfare economist may not be bothered by this objection, though. She maintains that we can always seek pareto improvements once the initial distribution has been decided on. If we are prevented from attaining an equal distribution of goods

because of an unfair initial distribution, it still makes sense to make the best of the situation and to seek pareto optimality. Her claim is that pareto optimality is *rationally* required from any initial point.

The claim that any pareto superior move is rationally required can also be disputed, though, when one considers a dynamic model. There are situations in which a rational individual might not want society to make a pareto improvement for some strategic reason. Suppose, for example, social policies are decided on by majority rule voting. Suppose that the pareto improvement under consideration does not benefit (and, of course, does not harm) Theo, and that Theo believes that if the society does not make this pareto improvement it will consider another one which does benefit him. Then he may oppose this one in the hope that the next one will be made instead. We can show that this situation can lead a society to reject all pareto improvements. Imagine that the society is a democracy of three members, Theo, Saundra, and Denise, and policies are decided on by voting for each policy in turn. Suppose that the policy being considered is a pareto improvement that benefits Saundra, but leaves Theo and Denise the same. In Table 1, A is such a policy. If Theo and Denise think that they will be able to replace it with a policy that benefits them, for example, B or C, respectively, then they will vote against the policy under consideration. The three of them might, for this reason, be unable to come to agreement on any pareto improvement at all, since any two of them could be looking ahead to another policy that would benefit them more, and hence each policy is defeated by a vote of 2 to 1.

Table 1

	Initial allocation	Policy A	Policy B	Policy C
Theo	100	100	200	100
Saundra	100	200	100	100
Denise	100	100	100	200

Furthermore, the assumptions of the PCM rule out important aspects of strategic behavior, and thus put in question the rationality recommendations derived from the PCM model. The PCM guarantees that there are no opportunities for strategic behavior by precluding force and fraud, and by assuming away opportunities for collusion. But this means that agents will not exercise strategic rationality, either because they may not or because the opportunities don't arise. In a theory of justice, the claim that force and fraud must be precluded (and what that amounts to) should be a theorem, not a postulate. This is especially true if the theory purports to derive the constraints of justice from rationality, as it is in the welfare economics argument, since agents will find this restriction arbitrary from the point of view of rationality. Otherwise, force and fraud have to be ruled out by a prior moral claim, making morality theoretically prior to justice. If we drop the assumptions that rule out strategic rationality, then there is no guarantee that pareto optimality will arise from free trade, since individuals may be defrauded to make trades that they would not want to make were they not defrauded. The no-collusion idealization of the PCM (C3) is also a problem for the rationality argument, since in a model that resembles ours in the sense that there are many opportunities for collusion, one cannot show that pareto optimality arises from free trade. Thus, for the purposes of defending the pareto criterion on grounds of rationality, the PCM is not robust.

David Gauthier, in *Morals By Agreement*,[14] defends pareto optimality as a criterion of impartiality in the PCM. Gauthier argues that the PCM is a "morally free zone," that in it morality is neither necessary nor possible. Morality for Gauthier consists in the constraints individuals agree to place on their pursuit of unbridled self-interest, in order to make possible cooperation for mutual advantage. In the PCM, however, there is no need for cooperation to achieve the greatest possible social and individual welfare; the competitive outcome is pareto optimal. Thus, he claims, there is no need for morality.[15] And since any changes in the distribution of goods from the competitive outcome would make some worse off, it is not possible to come to mutual agreement on any constraints on the pursuit of self-interest. So

morality is also impossible in the PCM. But there is nonetheless a moral quality that Gauthier points to, and that is that the PCM is *impartial.* A process is impartial if it does not unfairly benefit or burden anyone. Gauthier's claim is that the market confers benefits and burdens only on those who freely contract for them, to the extent to which they contract for them.

Gauthier argues that three features constitute the impartiality of the PCM: (1) there is free activity—individuals enter into only those interactions that they choose; (2) there are no externalities; and (3) the outcome is pareto optimal.[16] It is immediately clear why the first two should be criteria of impartiality: without free activity some could be forced to conform their behavior to others' wishes; and if there were externalities, then someone's endowment would be seized or damaged without her permission, or someone would benefit without sharing the cost of production of the benefit. In either case some would be burdened for the benefit of others. Gauthier argues that pareto optimality is also a criterion of impartiality with the following ingenious argument. In a PCM the unique pareto optimal outcome from that initial distribution is the only possible one in which everyone gets what, and only what, she pays for. We can see this by imagining that we enforce a move from a pareto optimal outcome to another outcome that does not worsen everyone. It is worth quoting Gauthier at length:

> [E]very alternative [to the competitive outcome] must involve a diminution in some person's utility, and we may treat this as either a loss of benefits or an increase in costs. If the former, then the person will not receive some benefit for which in free interaction she paid the cost; if the latter, then the person must pay some costs in addition to those that in free interaction were sufficient to cover all of her benefits. In either case her interests are prejudicially affected, and she may reasonably complain, if someone else enjoys an increase in benefits or a reduction in costs, that the other person has benefitted at her expense.[17]

So from the perspective of the competitive outcome, any other state would penalize some, possibly for the benefit of others, and thus would be partial.[18]

If impartiality is required by justice,[19] Gauthier's argument would seem to show that pareto optimality is a criterion of justice. That is not to say that pareto optimality guarantees justice, but that it is a necessary condition of justice. Another necessary condition

of justice in the market is that the initial endowments of the agents are just. Given a just starting point, an impartial process preserves justce.

Gauthier makes this argument only for the PCM, however. Can the argument be extended to the real world in which the conditions of the PCM don't hold? Not straightforwardly, because there is no pareto optimal point from which to begin—outside the PCM there may not be an achievable pareto optimal outcome—or to get to. Yet his argument seems to depend on there being an anchor point from which the alternative unjustly penalizes someone. What if, instead of forecasting a pareto optimal state, we look at each potential pareto improvement in turn and apply Gauthier's argument to isolated pareto improvements? That is, suppose that we begin with state A, and A′ offers some pareto improvement. Can we give an argument similar to the above to show that justice requires that we make it?

Again the answer is no. Recall the argument to show that pareto improvements are not rationally required. If some agents foresee better opportunities in the future by rejecting a proposed pareto improvement, they may rationally refuse it. The difference between this case and the PCM is that given an initial starting point there is only one pareto optimal point that is possible in the PCM, but in this case we are considering only pareto improvements, of which there may be many, each with different beneficiaries, and each with uncertain paths into the future. There can be no guarantee that all would agree on any given pareto improvement. Equally, they may refuse it on other moral grounds. For it may be the case that by refusing A′ a new opportunity A′′ arises which benefits more (though maybe not the same) individuals or individuals who are somehow more deserving. Thus the impartiality argument from the PCM is also not robust to relaxing its assumptions.

2. Arguments for Pareto Optimality from the OP

Contractarians take justice to be the outcome of a pre-cooperative situation in which the guidelines for cooperative social interaction

are agreed upon by more or less rational individuals. Contractarian theories try to avoid, as much as possible, imposing a conception of the social good, apart from weak assumptions about what the individuals in the pre-cooperative state can be supposed to want, and allow the citizens of the state to come to agreement on constraints on their individual pursuit of the good. The contractarian test of principles of justice is not only whether the individuals would agree to them in the initial situation, but also whether they remain motivated to comply with the principles in the future, and whether the social order that arises from that compliance is stable.[20, 21]

In this section I examine three contractarian arguments that bear on the pareto criterion and that begin from the Rawlsian OP. I am not attempting to provide a general critique of Rawls's *Theory of Justice* here, however. My aim is simply to examine arguments for pareto optimality as a criterion of justice that arise from deliberation about instrumental rationality in the Rawlsian OP. The first argument I shall examine is Rawls's lexical maximin argument for the difference principle and the second is an argument developed by Prakash Shenoy and Rex Martin[22] for a Rawlsian conclusion on the pareto criterion, the "collective asset argument." Finally I consider an argument of Gauthier from the conditions on individual rationality in the OP.

In Rawls's *A Theory of Justice*[23] the original position consists of equal, rational individuals behind a veil of ignorance who choose the principles of justice to govern the basic structure of society. The veil shields them from knowledge of their particular preferences, talents, attributes, and social positions. Since they know nothing of themselves which would differentiate them as individuals, and they are equally rational, there is no room for disagreement on the principles of justice; what will seem agreeable to one will be agreeable to all. Rawls argues that in this ideal situation the familiar two principles of justice, lexically ordered, would be chosen:

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.[24]

Since the ordering is lexical, Rawls gives higher priority to concerns of "equal basic liberties." But in the second principle Rawls makes distribution the other basic concern of justice. The first thing to notice is that inequalities are specifically allowed by part (a) of the second principle, usually called the difference principle. Rawls claims both that the principles of justice are compatible with efficiency, and that justice is prior to efficiency.

[I]n justice as fairness the principles of justice are prior to considerations of efficiency and therefore, roughly speaking, the interior points that represent just distributions will generally be preferred to efficient points which represent unjust distributions.[25]

To illustrate, suppose that we have a society of two individuals, i and j, in state A, and consider three other possible social states, B, C, and D, with cardinal utility profiles[26] as follows.

Table 2

	A	B	C	D
i	1	2	2	3
j	2	2	3	100

State B is a pareto improvement on A, C is a pareto improvement on B, and D is a pareto improvement on C. Rawls (most plausibly) uses the lexical maximin principle which requires that once the position of the worst off has been maximized, the second worst off member's position should be maximized, and so on until the best off member's position is maximized.[27] Using this principle to make the argument, (which I will call the "lexical maximin argument"), the difference principle is amended to read "to everyone's advantage," and it is sometimes referred to as the lexical

difference principle.[28] The lexical difference principle chooses D in the example above; this way Rawls can validate his claim that:

> The problem is to choose between [the many efficient arrangements of the basic structure], to find a conception of justice that singles out one of these efficient distributions as also just. If we succeed in this, we shall have gone beyond mere efficiency yet in a way compatible with it.[29]

But if the lexical maximin argument is taken as the argument for the two principles, then justice is not simply "prior to" efficiency, rather it *requires* efficiency. Equality, for instance, is shown by this example not to be a criterion of justice if D is chosen over C.

Rawls's lexical maximin argument relies crucially on the veil of ignorance of the original position. Lexical maximin is only a plausible choice rule if there is no good basis for estimating the probabilities of future possible states. One unassailable principle of rational decision is that one ought to use all of the information one has available, if only by choosing to ignore it because of time or calculation constraints. When good probability estimates of the future are available, expected utility maximization, weighted by one's risk posture, is a more reasonable choice rule than lexical maximin. But in any actual situation in which the principles of justice are to be applied, there is more information available about future possible states. So the plausibility of lexical maximin relies on the unique ignorance of persons in the OP.

The contractarian must also convince us that rational individuals will continue to comply with the agreement in society. If individuals feel that, knowing now what the outcome of their agreement is, they could make a better agreement if a renegotiation were forced, then they would press for renegotiation, and thus fail to comply with the rules based on the original two principles.[30] Since the plausibility of the choice rule in the OP depends on its artificially imposed ignorance, people outside the OP will not be motivated to comply if they feel they could have done better had they known more about themselves. Thus compliance with the two principles is questionable even if they do agree to the principles in the OP. Compliance and agreement are symmetric. Seeing that they would not be inclined to comply with the agreement, the individuals in the OP would not be inclined to make it on the basis

of lexical maximin, indeed they may even agree to wait until they have more information before agreeing on principles of justice. The problem I find with the OP, then, is that by changing a few assumptions slightly, for example, by letting the agents know what their special talents are, the argument for the lexical difference principle fails. In this sense, then, the argument fails to be robust: it is a theoretical failure of the idealization. That differs from the failure of robustness in the case of the PCM, which is an empirical failure. Nonetheless, the Rawlsian OP is not a robust contractarian model for the lexical maximin argument for the pareto criterion.[31]

Lexical maximin is not the only possible argument for the difference principle, and so also for the pareto criterion. Shenoy and Martin[32] prove that if chain connection[33] is supposed, then the difference principle is equivalent to choosing the pareto optimal states of affairs from among a set of possible states of affairs, and then choosing from this smaller set the state of affairs that is most egalitarian. Martin[34] argues that the pareto optimality half of this principle follows from Rawls's idea of collective assets; I shall call this the "collective asset argument." This argument supports pareto optimality only as the third criterion to be applied, and then only when chain connection holds.[35] Recall that for Rawls the principles of justice are lexically ordered, which means that the second principle is to be applied only to the set of distributions which meets the criteria in the first principle, which requires equal basic liberties. The second principle is then composed of two parts lexically ordered. On Martin's interpretation, first fair equality of opportunity should be secured and only then should the difference principle be applied. So the collective asset argument is designed to support pareto optimality only from the feasible set whittled down by both the first principle and the principle of fair equality of opportunity. These principles may constrict the set the difference principle is applied to in one of two ways: it may leave the pareto frontier untouched at some points, removing only those points that represent gross inequalities, as in Figure 1a, or it may shift the feasible set in, precluding all pareto optimal outcomes, as in Figure 1b.

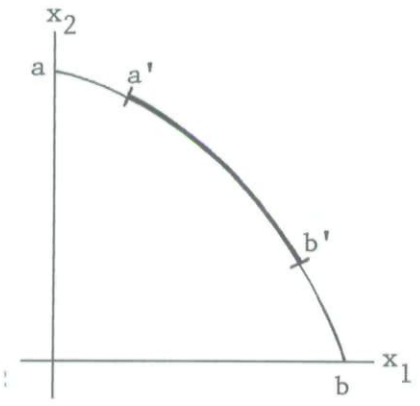

Figure 1a

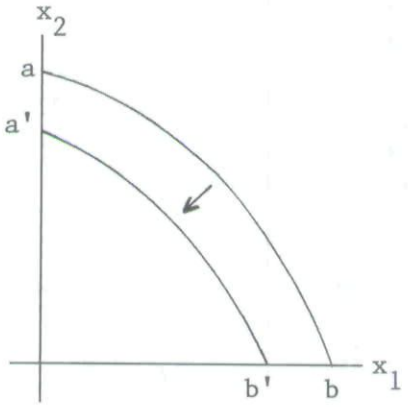

Figure 1b

If chain connection holds, that is, if the prospects of all are rising whenever the prospects of the best off and the worst off are, then we have a case like that in Figure 1a. Shenoy and Martin claim that their pareto optimality-egalitarian criterion is equivalent to the difference principle only in those cases where chain connection holds. And so I shall evaluate the collective assets argument assuming chain connection.

Rawls argues that talents and abilities which one is born with are morally arbitrary, or undeserved. The distribution of natural assets is a natural fact, but what difference they make to individuals is determined by social institutions. This is true in two senses. Which talents are valued depends on whether they can ever come to light or be rewarded; being able to jump very well and having great eye-hand coordination was not very useful in Victorian England, though now in the U.S. men with such abilities can become rich and famous in the NBA. Secondly, how the rewards to the talents are distributed depends on social institutions. The idea of justice as fairness is that "men agree to share one another's fate. In designing institutions they undertake to avail themselves of the accidents of nature and social circumstance only when doing so is for the common benefit."[36] Hence natural assets should be used as collective assets for the benefit of all. This principle of mutual benefit may be understood as claiming that the assets ought to be used to benefit the greatest number. Normally this means that all continuously improve their position. But if only

one portion of society can benefit, then the asset ought to be used for their benefit rather than wasted, so long as no one is worsened. Now if we can make a change that benefits some without harming anyone, then by the mutual benefit principle the change ought to be made, unless there is a better change that benefits more. This shows that all pareto non-optimal states (that fall within the constraints imposed by the lexically prior principles of justice) are inferior to some pareto optimal states, so none of the non-optimal states ought to be chosen.

The main objection to pareto optimality is the objection from envy.[37] The objection from envy is an individual's objection to the effect that he does not want others to be better off than he. Rawls rules out reasoning from envy in the OP for two reasons. The moral reason is that envy is "generally regarded as something to be avoided and feared"[38] and so ought not to count as a reason for decisions in the OP. Second, envy is a special emotion that is not always generated, and so Rawls assumes it away for the sake of simplicity. Nonetheless, he recognizes that it could become a destabilizing force in society, and thus he needs to examine the society that is ordered by the two principles to see if they are subject to envy. The kind of envy that is dangerous, Rawls argues, is "the propensity to view with hostility the greater good of others even though their being more fortunate than we are does not detract from our advantages."[39] Rawls allows that envy is a legitimate concern of justice when inequalities are so great that they undermine individuals' bases for self-respect. But, he argues, such envy doesn't arise in the well-ordered society based on the two principles of justice, since it is designed specifically to further the bases for individuals' self-respect. Hence it need not be a concern for the individuals in the OP.

Rawls's argument that destructive envy will not arise in the society ordered by the two principles is quite plausible. However, the flip side of envy, which is jealousy, cannot be so easily disposed of, and will, I think, particularly affect the collective assets idea. Jealousy occurs when one has possession of a relatively rare good that one does not want to share, or others to get more of. Jealousy is dangerous for Rawls's two principles if it causes individuals to renegotiate outside the OP. Talents and attributes exist attached to

individuals; they do not exist in a storehouse to be divided up among all members of society, although it could easily seem that way to those in the OP. Individuals who possess talents and attributes as part of their hands, bodies, brains want to benefit from them, and will regard them jealously the more that they must use them for others' benefit. Since they know that in a system of natural liberty they would benefit, those who have talents will be motivated to renegotiate. The veil of ignorance in the OP makes it impossible for individuals to guard against this, since ignorance of their talents persuades them to share the assets collectively. In other words, they agree out of aversion to risk. But if individuals will not subsequently comply with (legislation constructed against a background constitution based on) the two principles, the response to envy also fails, since it relies on the OP and the two principles. The veil of ignorance makes the two principles plausible in the OP, but it is an artificial device which, when drawn aside as it must be in society, takes with it the appeal of the principles.[40] Thus the OP argument for the pareto criterion is not robust to realistic adjustments.[41]

Gauthier, in "Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework,"[42] argues that in the OP pareto optimality is a condition on individual rationality for the kind of individuals we find in the OP, what he calls "everyman." Given the information conditions of the OP, and the principle of rationality that each maximizes her expected outcome, everyman wants to maximize the expectation of each, but since he could be any of these, he wants to maximize the expectation of any one individual only as long as the expectations of each other individual are not lowered. Hence it is rational as an everyman in the OP to agree only to those principles of justice which guarantee that there is no way that one individual's expectation could be increased without lowering that of any other by rejecting those principles of justice. But this just is the pareto criterion. Again, however, this reasoning only makes sense within the OP, where one's sense of an individual's interests is limited. Individuals do have interests that can oppose pareto optimality even if they are not envious: an individual rationally prefers a pareto non-optimal situation where he does better to a pareto optimal situation where

he does worse. The OP gives individuals a kind of anonymity that they might rationally reject outside it.

3. Pareto Optimality as a Regulative Ideal of Justice

We have seen that several crucial problems arise in the real world for the arguments for the claim that pareto optimality is a criterion of justice. The PCM arguments are ruled out because the opportunities for force and fraud and strategic behavior on the part of real individuals derail the argument that guarantees that a pareto optimal outcome will arise in the free market. Gauthier's argument from the PCM model founders when we relax the assumption that information about the future is perfect or that we can always forecast a static pareto optimal state. The Rawlsian arguments are ruled out because the principles of justice fail the compliance test outside the veil of ignorance imposed in the OP, and thus would not be chosen to begin with.

In this section I shall attempt to give a different contractarian argument for the claim that pareto optimality is a condition of justice that avoids the problems with the arguments I have examined. In particular, then, the argument must not rely on the PCM, and it must solve the compliance problem; it must allow for the possibility of change in the future, and must not require that there be no externalities.

To begin, I imagine a situation in which individuals know their talents, attributes, and limitations, and although they make reasonable estimates about their possible futures, they may disagree with one another about the feasibility of future states. I will imagine that individuals are rational, that they want to further their own best interests, and that they appreciate the fact that society can be, in the Rawlsian phrase, "a cooperative venture for mutual advantage." Their primary motivation to justice is the cooperative surplus, but they are also capable of a sense of justice, so that they also develop an affective attachment to societies and principles they believe to be in the best interest of the long-term cooperative enterprise. In this environment, just policy decisions are those decisions on which all would agree and with which all

would comply. Gauthier, in *Morals By Agreement,* provided an argument for the existence of moral constraints which would meet these requirements, and I shall assume that they exist in roughly that form for this paper. The question to be asked now is: will pareto optimality be a condition of just policies in such a society?

In this environment there is not enough information about the future to specify the pareto optimal states, and since we are not assuming away the possibility of force or fraud or collusion, it is not clear that pareto optimality would arise from apparently voluntary transactions anyway. The imperfections in the model force us to consider choices among alternative social policies, some of which offer pareto improvements over the status quo. Since finding the pareto optimal states of affairs assumes an artificial static endpoint, I shall consider the pareto principle to require simply that some pareto improvement be made whenever one can be.

First I want to establish that, facing a set of pareto improving policies, the contractarian notion of justice does not allow every pareto improvement to be refused. To see this let us consider the process of coming to agreement on social policies that affect utility distribution. Two questions have to be answered to begin with: (1) how is the decision to be made? (2) how are the gains from the policy to be divided? The two are interconnected in any society concerned with justice, since agreement on the first point will depend on agreement on the second. If the gains are not justly distributed, then the policy decision is not necessarily just either. Further, a third question arises in considering the process of coming to agreement on the second: (3) is envy to count as a reason for refusing agreement?

We can begin by showing that envy is not a rational objection to a proposed policy, and hence will not be used by rational agents. An objection is envious if the objector raises it only to block gains by others. To allow envy to count is to jeopardize all agreements that do not equally benefit all persons. Then any person can expect that many policies that would benefit her, even if they harm no one, would be refused, and this is something no one would want. If one policy is refused because it leads to envy, it may engender a spiteful response (i.e., someone might respond by pretending

envy even when they don't feel it in order to "get back at" someone who previously objected enviously) which might then be used as a reason to refuse another policy. Thus a just society, one that wants agreements that will be complied with, will agree not to count each others' envy as reasons.[43] Now one might object that the rationality of ruling out envy depends on how likely it is that one expects to be on the winning side rather than the losing one. But I think not. Only the person in the worst position could be unconcerned by the prospect of another's envy ruling out one's best social policies, and so find it rational not to rule out objections from envy. If one ever foresees being in a position other than that bottom position, then one would be concerned about it. It seems unreasonable to suppose that anyone would ever be in this position.[44] Agents will comply with this agreement as long as they foresee future interactions in which they benefit differentially.

In a contractarian theory, the answer to (1), how the decision is to be made, is that all must agree. Since our interest concerns the pareto principle, we want to know whether rational individuals agree to make pareto improvements. One difficulty arises when there are several pareto improvements from which to choose, or different routes to end-states that are pareto non-comparable and have different implications for different individuals. Let us look at an example, illustrated in Table 3 below. Suppose that the society is at state A and may choose policies that lead to development B and then B′, or C followed by C′, or D followed by D′. If it were considered as the only alternative, B would be pareto optimal. Since they consider no alternative that anyone likes better and no one argues from envy, B is chosen. However, as in the earlier Theo-Saundra-Denise example of Table 1, when there are other possible alternatives with different implications for the individuals, the situation changes. How are just societies to evaluate these possibilities? First we can rule out the status quo (A). For to remain at the status quo is in the first period to penalize one of the two individuals, and in the second period it would penalize both. We can also rule out development B followed by B′, since the ultimate state, B′, is pareto inferior to the ultimate state in either of the other two development routes, which for this case are pareto optimal, and there being no envy would be chosen

over B′. Between C followed by C′ and D followed by D′ the pareto principle makes no choice, for C′ and D′ are pareto non-comparable, and so, one might argue, CC′ and DD′ taken together are as well. The choice of the two is a bargaining problem, to be solved by other principles of justice, or principles of rationality in bargaining.

Table 3

	A	B	B′	C	C′	D	D′
Bob	1	1	2	3	3	2	2
Sarah	1	2	2	2	2	2	3

There are three important differences between our analysis in this case and that of the Theo-Saundra-Denise case. First, in this case we look beyond the first changes forged by the policies to future developments from those changes. This causes Sarah to prefer D to B and C, even though in the first period of the policy implementation there is no difference for her in the three. Second, we have ruled out envy explicitly. This means that between the B, C, and D possibilities, C and D are the ones to be considered. But at this point we would face the same problem as before, were it not for the final difference in the examples. Third, we allow the parties to bargain, rather than only to vote.

This argument can be extended by induction to any number of development steps into the future to show that if all possible improvements are considered together, not all pareto improvements may be refused. Considered together, no individual has an incentive to refuse all of them, since none make him worse off and some may benefit him. And he has no reason to refuse all of them, since he cannot, by hypothesis, hold out for a better one, and envy cannot count as a just reason. While this does not show us which improvement should be chosen, it does suggest that

justice requires that some improvement be made when a pareto improvement is possible.[45]

The only uncertainty now is over which pareto improvement to make, or question (2) above. Can any answer to this question lead us to revise our answer to question (1), that is, is there a compliance problem, consideration of which leads us to reject the pareto principle? Yes. The problem of choosing the particular future state to implement is a social choice problem that must take distribution into account in the very making of the agreement; agreement on pareto improvements hangs on the distributions they provide. There seem to be three ways to make the choice: a preset rule, such as Rawls's difference principle, or a voting rule, or bargaining. As long as one of these will guarantee compliance, the pareto principle survives. Voting rules are subject to paradoxes and agenda setting.[46] Preset rules are likely to be violated by predetermined losers, as I argued will result from the collective assets argument. Bargaining takes each individual's concerns into account by letting each represent herself in each policy decision situation.[47] Each wants to settle but can only offer so much to the social whole for a settlement on his favorite. The one who gains most will be able to redistribute most to buy agreement from others. Since agreement has to be unanimous, and all want an agreement that will be complied with, there is no incentive to agree and then renege. One who would renege would rather hold out on the agreement, since reneging damages future prospects for agreement. Yet no one holds out forever, since that too damages future prospects for agreement, and amounts to rejection of the agreement out of envy.[48] Acceptance of any agreement at all requires that all are reasonably sure that it will command compliance.

One might argue that I have simply given an argument for bargaining over voting here. That response seems partially right; my argument is an argument for bargaining over voting. Bargaining provides the way to split up the social product without having to sacrifice pareto optimality, allowing each person to make a claim, an argument, for herself. The great advantage of bargaining is that it allows one to divide the whole social product; in bargaining, rationality requires that the outcome be pareto

optimal. So if the argument is an argument for bargaining, then it is an argument for pareto optimality as a requirement of rationality, as well.

Notice that to secure the agreement and compliance through bargaining in a dynamic world, society must take a broad view of potential pareto improvements, so that no one's potentially best pareto improvements are left unconsidered. To assure compliance, the outcome must be rational for each and it must begin from an acceptable bargaining position,[49] and being rational for each, the bargain is just. If there were any pareto improvements yet to be made, someone would have reason not to comply, and having reason not to comply, others would not agree. Thus the outcome of the bargain may not be pareto inferior to any other.

On the contractarian view, justice is forged through agreement on the social structure, in which each individual seeks his own best outcome, but all are concerned to achieve a pareto optimum where possible, or to capture the optimal combination of pareto improvements. Just persons must be willing to entertain dialogue about pareto improvements, and to bargain with others over the gains from those improvements. For without this there is no compliance, and there being no compliance there would be no agreement. Considered all together, everyone wants agreement on some pareto improvement, yet all realize that agreement on any depends on the right compromise on distribution. I shall call someone who holds the view that pareto optimality is a criterion of a contractarian theory of justice a *paretian contractarian*.

There remain two troubling objections for the paretian contractarian. Both objections claim that information requirements for using the pareto principle to make social decisions in the way I have advocated are too extensive for it to be useful in the real world. One concerns the difficulty of predicting the course of future development, and the other concerns the information that must be gathered about individual preferences in different social states. Regarding the first problem, one needs to know all the possible pareto improvements that can be made now, and the possible ramifications for the future that can be expected from each of them. Clearly this is a strict requirement. Without this information, though, unexpected possibilities could

arise which cause individuals to renege on their agreements, seeing that their interests were not fully considered in the original agreement.

The paretian contractarian might respond in one of two ways. First he may argue that pareto improvements of any kind are valuable, so even if a potential pareto improvement is overlooked, the one that is made is not against justice. But this response will not do because the argument that the bargain would be complied with depended on each person's favorite pareto improvement being recognized and considered in the bargain. If an individual believes that she might do better by holding out for some as yet unrecognized improvement, then compliance breaks down. Second he may respond that all possible pareto improvements, discounted by their probability of successful achievement, are to be considered in the bargain. But unless all agents agree on the possible improvements and their probabilities, compliance and agreement break down once again.

The other information difficulty is how to obtain the preferences of individuals for different social states. In order to compare social states, individuals have to know what their welfare is in those states. Their welfare can be determined objectively or subjectively. I have assumed a subjective conception of utility in this paper, and would argue that it is the only one that is compatible with contractarianism; an objective assessment of welfare would not ensure compliance. But to make subjective assessments, individuals have to know what they would prefer if they were the persons in the alternative states. If the preferences are themselves state-dependent, then this is a serious difficulty in forecasting. What happens if they are wrong, for example, if an individual supposes she would be very happy in a state in which she turns out to be miserable? Again, compliance would be threatened if that state came about, and knowing this is possible, compliance and agreement are threatened in the present contemplation of the future state.

I have argued that at least three epistemological difficulties become very important for questions of justice: the difficulty of defining the pareto frontier, the difficulty of imagining possible futures and their consequences for the distribution of goods, and

the difficulty of projecting one's preference orderings in different future possible states. Added to these is the difficulty in making these items of knowledge *common* knowledge in a society, as they must be for effective and fair bargaining. These difficulties surely are, in a precise sense, insuperable for human minds. On the other hand, they are, under other descriptions, the kinds of things that ordinary conversations and political debates consider all the time. Justice requires that we come to some agreement on social states, and if this comes about as a kind of bargain, then what drives agreement is the belief that each is making in some sense an equal sacrifice of the cooperative surplus in order to get agreement.[50] The only way that the belief can be generated and shared is if the bargain takes place by means of shared public discussions and democratic institutions, in which reasons are given and asked for against a set of shared norms for justification, and in which each feels her voice is heard. So in seeking general criteria of justice, an examination of pareto optimality suggests that we ought to choose a just democratic process as much as a just outcome criterion.[51]

There is another reason to believe that the process of coming to agreement is as important for justice as securing the pareto frontier. Individuals are sometimes not rational, and so we might imagine that sometimes they will not agree to even an obvious pareto improvement. If we were to say that pareto optimality is a necessary criterion of justice, then should it not be imposed on them? Since no one is in the epistemic situation to know that a policy guarantees pareto optimality now and into the future, there is no one who can justifiably impose such a change. So it is better to invest in a process that is likely to bring about agreement and compliance, that often results in mutual gains, and that avoids taking account of envy. But the process cannot be chosen outside of all considerations of the content of the agreement either. In order to secure agreement, those best able to stall agreement have to be content with their share. And this means that the content of the agreement is as important to securing it as the process is.

The objections presented in this section show that the pareto criterion cannot be an operative criterion for societies. It is not possible to project a determinate pareto optimal state, and even

less is it possible for persons to know all possible pareto improvements or even their preferences over them. But the argument against using envy as a legitimate objection to proposed social arrangements suggests that the pareto criterion is a plausible regulative ideal. That is, if we could have the ideal information about society that we would need in order to know what improvements are possible, and we could know our preferences over them, and we had common knowledge of these facts, then the pareto principle would be justifiable. Lacking such knowledge, however, we can glean from this discussion of the problems that arise in the search for just social policies that it must take into account the possible futures that all individuals look to in formulating their preferences and concerns. Though I began with a question that apparently concerned static distributions of goods, I have argued that in a dynamic world we cannot come to a firm conclusion about the relation of pareto optimality to justice in distribution. However, we are able to draw conclusions about the *process* of making just distributive policies. If there is to be agreement at all, individuals have to bargain over the possible futures in good faith, making reasonable judgments about the consequences and feasibility of each others' visions of the future, and engaging in give and take for the sake of cooperative agreements in the future. This means that for the sake of agreement, and so justice, individuals will have to participate in policy formation. They will have to insure that their visions of the future will be heard and evidence and argument for it presented. And finally this means that individuals will have to be ready to consider futures that may be contrary to their traditions and prejudices and apparent immediate self-interest for the sake of cooperation, that is to say, for the sake of their enlightened self-interest in the face of an uncertain future.[52]

Notes

1. This paper will not consider whether the "distributive paradigm" of justice, which takes social justice to depend largely on the distribution of goods in society, is the appropriate framework for discussions of justice. I am going

to be working within that paradigm. However, Iris Young, *Justice and the Politics of Difference* (Princeton, N.J.: Princeton University Press, 1990), provides an interesting critique of the distributive paradigm. See esp. chap. 1. I shall respond simply that distribution of goods is surely part of justice, though surely also not the whole of it.

2. But see Julian Le Grand, "Equity versus Efficiency: The Elusive Trade-off," *Ethics* 100 (1990): 554-68 for a discussion of some problems with identifying pareto optimality and efficiency.
3. Vilfredo Pareto, *Manual of Political Economy* (New York: Austus M. Kelley, 1971), chap. 6.
4. I am using "welfarist" in the sense of Amartya Sen. See *On Ethics and Economics* (Oxford: Basil Blackwell, 1987), which refers to theories which take the justice or goodness of a society to be completely determined by the utilities of the individuals in it.
5. An example of a welfare economist using pareto optimality as a normative criterion is Gordon Tullock, "Inheritance Justified," *Journal of Law and Economics* 13 (1970): 465-74.
6. These criteria are redundant to some extent. In particular, C1, C3, C4, C5 would be sufficient to characterize the PCM, properly understood. I add the others just to make the conditions clear.
7. See Richard Thaler, "Psychology of Choice and the Assumptions of Economics," in Alvin Roth, *Laboratory Experimentation in Economics* (Cambridge: Cambridge University Press, 1987), pp. 99-130, for a review of experimental evidence of violations of economic assumptions about human behavior.
8. Recall that there are no transaction costs, so that if a person is indifferent to a trade there is no sense in which he would prefer not to trade; we shall suppose that in these cases the indifferent party makes the trade.
9. But see Allan Gibbard, "What's Morally Special About Free Exchange?" in *Ethics and Economics*, E.F. Paul, F.D. Miller, Jr., and J. Paul, eds. (Oxford: Oxford University Press, 1985), pp. 20-29, for a criticism of the notion that free exchange is prima facie a good thing.
10. Hal R. Varian, "Distributive Justice, Welfare Economics, and the Theory of Fairness," *Philosophy and Public Affairs* 4 (1974-75): 223-47. Reprinted in Hahn and Hollis, eds., *Philosophy and Economic Theory* (Oxford: Oxford University Press, 1979).
11. Building a lighthouse is rational whenever the expected value of the lighthouse is less than the expected cost of the lighthouse. For a large and profitable enough shipping company, that would be the case for some locations.
12. It is not quite a unanimous opinion in the economics literature that externalities harm third parties, that is, persons who are not party to trades. However, the argument that externalities are harmless depends on the assumption that costs of making voluntary agreements are completely

uniform across the market, and this is clearly violated in the world. See James M. Buchanan, "The Relevance of Pareto Optimality," *Journal of Conflict Resolution* 6 (1962): 341-54; esp. p. 349.

13. For a stimulating discussion of other problems with the welfare economics treatment of pareto optimality and justice, see Le Grand, op. cit.
14. David Gauthier, *Morals By Agreement* (Oxford: Oxford University Press, 1986) (hereafter "*MBA*").
15. Again, however, there must be no force or fraud in the PCM. Hence there must be some prior, apparently moral, constraint, contra Gauthier.
16. Daniel M. Hausman, "Are Markets Morally Free Zones?" *Philosophy and Public Affairs* 18 (1989): 317-33, interprets Gauthier as saying that impartiality is a result of the lack of externalities, not the pareto optimality of the market. But it seems clear from what Gauthier says at *MBA*, pp. 96-97, and from the way that he makes the argument for the impartiality of the market, as I show shortly, that he intends that pareto optimality is a condition of impartiality in the PCM.
17. *MBA*, p. 97. Note that he goes on to qualify the argument in one way: if the market outcome includes rent, which would happen if there were a fixed supply of a good lower than what would be demanded at the equilibrium price, then "optimality does not straightforwardly ensure that any alternative that was not worse for everyone would benefit some individuals at the expense of others" (p. 98).

    Hausman, op. cit., argues that rents can be generated in the PCM (p. 326) provided only that individuals have different levels of talents and/or endowments. He argues further that pecuniary externalities can arise in the PCM when technologies are allowed to change (pp. 328-29). Thus the PCM can only be morally free under very restrictive conditions, conditions which Hausman claims make the argument trivial. This argument does not affect the claim I am concemed with, that pareto optimality is a condition of impartiality. But I am essentially agreeing with Hausman's claim.
18. In private correspondence, David Gauthier tells me that this interpretation of his argument is not quite what he had intended. He points out that any pareto optimal alternative from the one in the PCM would also not be impartial, since it would benefit those for whom the alternative is preferable, at the expense of those for whom it is less preferred.
19. Recent work in ethics, especially by feminist ethicists, has given us reason to question the assumption that impartiality is a hallmark of justice. See Marilyn Friedman, *What Are Friends For?* (Ithaca: Cornell University Press, 1993), for an exemplary discussion of these issues.
20. While Gauthier emphasizes the compliance problem in his contractarian project, Rawls talks about the stability of the just society. This turns out to be more than a semantic difference. Gauthier appeals to instrumental rationality to construct his theory of justice, Rawls to a broader conception of rationality. I take it that if a theory of justice is instrumentally irrational in

the sense that it cannot solve the compliance problem, that theory is seriously flawed.

21. Social contracts are to be understood as *self enforcing agreements*, that is, as agreements which, once made, will be complied with because both the agreement and the subsequent compliance is rational.
22. Rex Martin, *Rawls and Rights* (Lawrence, Kan.: University Press of Kansas, 1985).
23. John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971) (hereafter "*TJ*").
24. *TJ*, p. 302. This is the "full statement" of the two principles. Martin, op. cit., presents the second principle ordered the other way, that is, with fair equality of opportunity coming before the difference principle. But this quote from Rawls matches, in the ordering of the principles, his presentation of the two principles on pp. 60, 83, and 250, as well.
25. *TJ*, p. 69.
26. There are two complicating features of Rawls's account that I am glossing over here. First, the two principles are to guide the design of the institutional framework and only indirectly the distribution of goods. But since it is the distribution of goods that rational individuals would look to in deciding on the principles, this simplification is generally a reasonable one. Second, Rawls would prefer to speak in terms of bundles of primary goods here, but the difference is not important for the point I am making here, provided that it is better from the point of view of justice either to allow individuals more of what they prefer or more of the primary goods.
27. The formal condition is nicely put by Shenoy in Martin, op. cit., p. 198, as follows. Let $A\varepsilon\{x=(x_1,...,x_n): x_1\geq x_2\geq...\geq x_n\}$ be the set of all feasible distributions, where x_i is the expectation of the ith most favored individual. Then the distribution $a^* = (a_1,...,a_n)$ is the one picked by the lexical maximin principle as follows:

    (i) $\max \{\min \{x_i:i=1,...,n\}:x\varepsilon A\} = a_n$

    (ii) $\max \{\min \{x_i:i=1,...,n-1\}:x\varepsilon A, x_n=a_n\} = a_{n-1}$

    . . .

    (n) $\max \{\min\{x_i:i=1\}:x\varepsilon A, x_n=a_n,...,x_2=a_2\} = a_1$

28. See *TJ*, pp. 81-83.
29. *TJ*, pp. 70-71.
30. Rawls relies on persons in the well ordered society forming a well developed sense of justice that gives them an interest in being just and encouraging justice to solve the stability problem, his analogue to the compliance problem. This cannot be a rational sentiment if my criticism of the OP holds.

I will present an argument shortly to suggest that there will be countervailing non-rational sentiments against the two principles in the form of jealousy.

31. The failure of Rawls's model is not the same as that of the PCM. The PCM purports to be an empirical idealization of life, while the OP is an idealization of a hypothetical initial state. I thank Geoff Sayre-McCord for pointing out that there are two senses of robustness in play here.
32. Martin, op. cit.
33. Chain connection holds just in case that whenever the prospects of the representative persons in the highest and lowest positions are both rising (or falling), so are the prospects of the representative persons of all other positions.
34. Martin, op. cit.
35. On Martin's view the difference principle is lexically ordered after the principle of fair equality of opportunity, and since this use of the collective asset argument is his, I follow him here.
36. *TJ*, p. 102.
37. One might also imagine an egalitarian objection which says that equality of distributive shares ought always to be preferred. But Rawls rules this out as something rational individuals who are concerned to further their own interests would not choose even in the OP, unless they are subject to great envy.
38. *TJ*, p. 530.
39. *TJ*, p. 532.
40. David Gauthier, "Justice and Natural Endowment: Toward a Critique of Rawls' Ideological Framework," *Social Theory and Practice* 3 (1974): 3-26, argues against the collective asset argument in a similar way, arguing that rational individuals will reassess their agreement once they discover their endowments outside the OP. However, his argument is not based on an emotion, jealousy, but rather on what he considers a principle of rationality:

> A man will not consider his agreement rational, if his share of [the social surplus beyond what could be obtained in "general egoism"] is less than that of some other person, unless (for no man is envious) greater equality could be achieved only by reducing the share of that other person, and not by increasing his own. (p. 16)

I would agree with this provided that an account of why rational persons forgo envy can be given. I attempt to give such an account in the next section. Without such an account the appeal to rationality simply assumes the pareto criterion.

41. See n. 31.
42. Gauthier (1974), op. cit.
43. Note the difference between my argument against envy and Rawls's—I argue that it is not rational for individuals to reason from their own envy, not that envy will not arise, or that it will not be dangerous.

44. I thank an anonymous referee for raising this objection.
45. An apparently difficult situation seems to arise when a society is faced with several choices that are pareto non-comparable to the status quo, as in the following example.

	A	B	C
Bob	1	10	98
Sarah	100	99	98

In this example all three choices are pareto non-comparable. Thus the pareto criterion cannot choose among these. In particular, it cannot be said that the pareto criterion sides with the status quo point simply because the other points are not pareto improvements, since the same can be said of each point given any of them as the status quo.

46. See for example, Charles R. Plott and Michael E. Levine, "A Model of Agenda Influence on Committee Decisions," *American Economic Review* 68 (1978): 146-60, and Robin Farquharson, *Theory of Voting* (New Haven: Yale University Press, 1969).
47. Not every bargaining structure will guarantee that each person's concerns receive equal consideration. I have in mind Gauthier's minimax relative concession bargaining solution, though this is not the place for a sustained defense of it. For the sake of the argument in the text I ask only that the reader suppose for the moment that there is a bargaining structure that formally treats each bargainer's concerns equally.
48. One might argue that bargaining solves compliance by definition, since the possibility of non-compliance is not a part of any formal bargaining model. But I am considering the possibility of using, say, force or fraud after the bargain is struck as a way of reneging on the bargain.
49. An acceptable initial bargaining position requires that claims to retributive justice have been settled.
50. I say "in some sense an equal sacrifice" because there are several ways of reckoning the sacrifice: one can compare the absolute amounts of the cooperative surplus beyond what one could secure oneself without cooperation, or the sacrifice relative to one's gain, for example. Which one we pick is surely an important matter, and will determine whether the agreement will be complied with. But it is also beyond the scope of this paper.
51. Gauthier provides an interesting discussion of how one might move from strategic bargaining to democratic politics in "Constituting Democracy," given as the 1989 E.H. Lindley Lecture, University of Kansas.
52. I would like to thank Neal Becker, David Gauthier, Rex Martin, Geoffrey Sayre-McCord, and an anonymous reviewer for very helpful comments on

an earlier draft. I gratefully acknowledge the University of Kansas for a grant from the General Research Fund, proposal #91-103, for work on this paper.

Ann E. Cudd
Department of Philosophy
University of Kansas
acudd@ukanaix.cc.ukans.edu