

The First Virtue and the Realistic Utopia

Jon Garthoff
Northwestern University

1. Introduction

John Rawls famously opens *A Theory of Justice* by characterizing justice as “the first virtue of social institutions”, explaining further that by this he means that “injustice is tolerable only when it is necessary to avoid an even greater injustice”.¹ This understanding of justice is reaffirmed later in the book, such as when Rawls asserts that principles of justice are “categorical imperatives ... in Kant’s sense”.² On this understanding, the subject matter of justice is right and wrong, permissibility and obligation, in the context of political relationships. Failing to satisfy this sort of prescription of justice not only fails to do what there is sufficient reason to do; more strongly, it violates or denigrates someone, or otherwise prevents individuals from standing in relationships of mutual respect with one another. Injustice so understood entails impermissibility and hence intolerability, and so on this understanding a theory of justice regulates societal action by prohibiting social structures that violate its all-things-considered prescriptions.

Rawls also sometimes characterizes theorizing about justice as a search for a “realistic utopia”; he expresses this understanding of justice most explicitly in his book *Justice as Fairness: A Restatement*.³ On this understanding, a theory of justice articulates a vision of society that is both better than the status quo and feasibly approachable through reform. This understanding of justice is also suggested by Rawls’s characterization of his work, at all stages of his career, as an exercise in “ideal theory”.⁴ Ideals regulate action differently from prohibitions. Ideals enlarge the imagination, so that people may envision new possibilities; they inspire people, transforming their beliefs about justice into motivation to act justly; and they provide clarity and focus, distilling the claims of a single domain of normative concern. But when a society fails to satisfy an ideal, and accordingly fails to achieve a realistic utopia, this does not as such entail that the prevailing social order is impermissible or intolerable.

These are distinct understandings of the subject matter of a theory of justice, and I believe there is a deep and never fully resolved tension in Rawls's thought between these understandings of his subject matter.⁵ This tension between viewing a theory of justice as a set of prohibitions against impermissible social structures – the “first virtue” understanding of justice – and viewing a theory of justice as an ideal – the “realistic utopia” understanding – arises in many contexts. Depending on which understanding of the subject matter of justice we attribute to Rawls, we arrive at distinct interpretations of many trademark Rawlsian positions.

One aim of this essay is to explain how this tension in Rawls's thought arises. A second aim is to articulate what I take to be the best Rawlsian response to the tension. This response makes use of Rawls's distinction between *constitutional essentials* and a *complete conception of distributive justice*.⁶ In Rawls's view, a society's constitution is adequate when all citizens are allocated an equal and adequate scheme of political liberties, all citizens have an adequate range of social and economic opportunities, and all citizens' humanitarian needs are met. In short, constitutional essentials encompass all of Rawls's conception of distributive justice except his second principle of justice, which subsumes his fair equality of opportunity principle and his difference principle. The second principle of justice governs the distribution of socially produced goods beyond the constitutional essentials.⁷

In this essay I argue that the best Rawlsian account of justice maintains that securing constitutional essentials has the normative force of prohibition, but that the fair equality of opportunity principle and the difference principle have the normative force only of a regulative ideal. On this view, a society's failure to realize constitutional essentials is an injustice in the “first virtue” sense; it is to be tolerated only where needed to forestall even greater injustice. A society's failure to satisfy the fair equality of opportunity principle or the difference principle, by contrast, entails as such only that the society is less than ideally just. This does not entail that an individual in that society is wronged by the prevailing social structure, nor does it entail that any pair of individuals in that society are unable to interact on terms of mutual respect. It does not

even entail that it is better, all things considered, to change the social structure so that it better satisfies the fair equality of opportunity principle or the difference principle. When justice is understood as a regulative ideal, it is possible for considerations of justice to be defeated by competing normative considerations. I do not claim that the Rawlsian conception of justice that emerges from this suggestion is Rawls's own view, but I believe it is highly appealing and that it deserves serious consideration.⁸

2. Justice as Fairness as Fully Directive

2.1 Optimization in Justice as Fairness

Rawls's conception of justice, which he calls "justice as fairness", consists in two distinct principles. The first governs the distribution of political liberties, and the second governs the distribution of other socially produced goods. He presents justice as fairness as follows:⁹

The First Principle of Justice

The Equal Liberties Principle. Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all.

The Second Principle of Justice

Social and economic inequalities are to satisfy two conditions:

The Fair Equality of Opportunity Principle. They are to be attached to offices and positions open to all under conditions of fair equality of opportunity.

The Difference Principle. They are to be to the greatest benefit of the least-advantaged members of society.

The second half of the second principle of justice, Rawls's difference principle, is an optimizing principle: it instructs us to organize society so as to optimize the amount of socially produced goods received by the group that receives the fewest such goods.¹⁰ Like any optimizing principle, the difference principle is, to use Seana Shiffrin's helpful term, *fully directive*.¹¹ excepting rare

instances of ties, the difference principle picks out a uniquely best social structure. In Parts 2 and 3 of this essay I discuss puzzles about justice as fairness that arise from the fact that it is a fully directive conception of justice.

I begin by registering some surprise that Rawls asserts a fully directive conception of justice, since he is sometimes claimed to have “slain the beast” of utilitarianism. This claim is hyperbole, to be sure; that beast clearly still stalks the philosophical landscape. But many prominent critiques of utilitarianism as a principle of individual action exploit the fact that the principle of utility is fully directive and thus appears to preclude individual commitments to particular values and appears to permit individuals no prerogatives to choose how to live.¹² It is reasonable to wonder whether analogues of these criticisms might carry over to the domain of political philosophy, and impugn the principle of utility – or any other fully directive candidate principle, such as the difference principle – as a principle of distributive justice.¹³

One such analogue would propose that societies themselves must be able to commit to particular values and must have prerogatives about how to act, and would infer as a consequence that societies cannot be subject to fully directive principles of distribution. I believe there is an interesting objection to the difference principle along these lines, but I will not present it here.¹⁴ Instead I focus on a different analogue to prominent objections to utilitarianism, one that objects more generally to the claim that a principle that purports to have the normative force of moral prohibition can be fully directive. This objection is that fully directive principles are too sensitive to capture the conditions under which social arrangements permit individuals to relate on terms of mutual respect. A fully directive principle of justice picks out a unique social arrangement – or an extremely small family of arrangements – as that under which individuals are able to stand in relationships of full respect. It entails further that even slight departures from this arrangement render it impossible, strictly speaking, for individuals to stand in fully mutually respectful relationships. Any fully directive account of distributive justice thus makes strong claims, and accordingly faces a high burden of argument.

In Part 3 below I develop this objection to fully directive principles of justice in the context of Rawls's theory.¹⁵ But first, in the remainder of Part 2, I consider two strategies for denying that justice as fairness is fully directive, and hence for claiming that justice as fairness is not subject to this line of criticism. Both of these strategies fail, however, and so in Part 3 I take on an alternative defense of Rawls's theory. This defense concedes that justice as fairness is fully directive, but denies it is problematic to assert a fully directive theory of distributive justice.

2.2 Institutional Division of Labor

On one strategy for denying that justice as fairness is fully directive, Rawls's claims about how the difference principle is to be implemented are claimed to prevent his theory from being fully directive. Rawls suggests that, rather than having each institution directly regulate its policies by the difference principle, we should instead have a functional division of labor among institutions.¹⁶ Thus the United States Department of Commerce may try to reduce unemployment in order to enhance economic growth and efficiency, for example, without considering directly the conditions under which reducing unemployment works to the benefit of the least-advantaged group; the Federal Trade Commission can likewise bust monopolies without aiming to do so in the way that would most benefit the least-advantaged group; and the Environmental Protection Agency can design policies for the preservation of wildlife and natural spaces without aiming directly at the benefit of the least-advantaged group. Rawls maintains that this division of labor among institutions is compatible with satisfaction of the difference principle because it is the entire structure of social institutions constituting the vehicle of societal cooperation – the “basic structure of society” – that is subject to the difference principle, not the component institutions considered severally.¹⁷ Rawls speculates that the best way for the basic structure to satisfy the difference principle is for it to contain a small number of institutions deputized as responsible for ensuring that the structure as a whole operates to the greatest benefit of the least-advantaged group. More specifically, he proposes that the institutions of taxation and transfer, such as the

Internal Revenue Service and the Social Security Administration, be charged with this function.¹⁸

Thus it might be thought that even though the basic structure as a whole must be responsive to a fully directive principle of distribution, the institutions that compose it need not be, and so their activities will not be straight-jacketed by such a principle.

There is considerable appeal in Rawls's proposed division of labor among institutions.¹⁹ If the several agencies each aim to benefit the least-advantaged group, there is a danger that the basic structure as a whole will act in an insufficiently unified and so self-defeating way. There is also a concern that the several agencies may lack information needed to exercise good judgment about which policies would work to the greatest benefit of the least-advantaged. I readily grant that considerations like these justify the division of labor among institutions, if we aim to satisfy the difference principle only approximately; and this aim is appropriate, if the principle has the normative force of a realistic utopia and so fails to override all competing considerations.

If the difference principle has the normative force of the first virtue of social institutions, however, then it is problematic to aim only at its approximate satisfaction.²⁰ Failure to satisfy the principle would then entail that some individuals are unable to stand in relationships of mutual respect. This would make it incumbent on us to investigate in detail how the institutions of the basic structure might each contribute to satisfaction of the difference principle, so as to improve on the extent to which the principle may be satisfied using the institutions of taxation and transfer alone. It would be a staggering surprise to discover that no efforts by any other institution could improve the condition of the members of the least-advantaged group.

Rawls can of course allow that we should reform the basic structure so that it achieves these more fine-grained satisfactions of the difference principle, if we have confidence about how to do it. His discussions of the point fail to reflect, however, the urgency that these matters have if we understand the difference principle to have the normative force of the first virtue of social institutions.²¹ And whatever Rawls's own views, the difference principle has much greater initial plausibility when understood only as a regulative ideal. As I noted earlier, and as I discuss at

greater length below in Part 3, it is difficult to believe a conception of moral standing so fine-grained that it demands complete satisfaction of the difference principle and asserts that no approximation to this is morally adequate.

2.3 Lexical Subordination

I now leave to the side Rawls's claim that we should implement the difference principle through a functional division of labor among agencies of government, and consider an alternative strategy for denying that justice as fairness is fully directive. This defense focuses on the lexical subordination of the second principle of justice to the first. The thought here is that since the second principle applies only in a context that assumes satisfactory constitutional essentials, including especially satisfaction of the equal liberties principle, the fully directive element of justice as fairness is contained.²² To understand how this defense works, it is helpful to recall Rawls's view that justice as fairness is to be satisfied through a "four-stage sequence".

At the first stage of this sequence, a conception of justice is selected. Rawls's argument from the original position inhabits this stage. In the original position, individuals lack access to information about themselves, such as their social class, race, ethnicity, sex, and natural talents. This information is excluded because Rawls maintains that it is irrelevant to the content of a conception of justice, and so maintains that access to it unfairly biases the selection of such a conception.²³ And this is not the only information excluded in the original position. Individuals in the original position also lack access to particular information about the society they live in, such as its natural resources, available technologies, and cultural composition; and the ground for exclusion is the same as the ground for the first, namely, this information is held to be irrelevant to the content of a conception of justice.

At the second stage, a constitution is selected in light of the conception of justice selected at the first stage.²⁴ The constitution is to secure the constitutional essentials, including especially satisfaction of the equal liberties principle. The constitution is also to be the principal vehicle of

societal decision-making. Particular information about individuals remains excluded at this stage, since it is irrelevant to selection of a constitution. Particular information about the society itself is *not* excluded, however, since constitutional design should be sensitive to this information. At the third stage, laws are selected in light of both the conception of justice already selected at the first stage and the constitution already selected at the second stage, subject to the same information constraints as in the second stage. Finally, at the fourth stage, laws are administered by judges, officials, and citizens, with no constraints whatsoever on available information.²⁵

The second principle of justice, including in particular the difference principle, is to be satisfied principally at the third stage of the sequence. Laws regarding taxation and transfer are selected at this stage.²⁶ This posteriority of the laws needed to satisfy the difference principle to the constitution needed to satisfy the equal liberties principle, the present proposal maintains, prevents justice as fairness from being fully directive, even though the difference principle itself *is* fully directive.²⁷ The difference principle's activity is constrained, on this proposal, by the other parts of justice as fairness to which it is subordinated. Thus even if the fully directive nature of the second principle severely restricts the choice of laws at the third stage, societies have latitude at the prior constitutional stage, which is not constrained by this principle.

This proposal fails, however, for it misunderstands the nature of lexical subordination. The subordination of the second principle of justice to the first entails that the difference principle applies only when the first principle is satisfied as well as it can be. It does *not* entail that the difference principle fails to apply to those institutions, such as the constitution, whose principal purpose is to satisfy the first principle. To see this, suppose a society enacts a constitution – call it Constitution X – that satisfies the equal liberties principle; and suppose further that this society enacts a set of legislative measures that best satisfies the second principle, in light of their prior choice of X. Suppose further that there is a rival constitution, Constitution Y, which would, in this society's historical circumstances, satisfy the equal liberties principle as well as X.²⁸ Now suppose next that a legislator in this society notices the following fact: had they adopted Y instead

of X, a set of legislative measures would be available that would benefit the least-advantaged group more than the actual set of measures – by hypothesis the best available – that were adopted under X. This legislator proposes to amend the constitution from X to Y, on the ground that this permits a set of laws that better satisfies the difference principle.

Given that all the legislators seek to implement justice as fairness, they should all endorse this constitutional amendment. By hypothesis X and Y satisfy the equal liberties principle equally well, so the lexical priority of that principle to the difference principle fails to show that this is a question to which the difference principle does not apply. If the difference principle decisively adjudicates among legislative measures under the hypothesis that a satisfactory constitution has already been adopted, then it also adjudicates among legislative measures that require constitutional amendment to be enacted. Notwithstanding its lexical subordination to the equal liberties principle, the second principle ramifies back up to the second stage of the four-stage sequence, and has normative force with respect to the selection of a constitution within the range allowed by satisfaction of the equal liberties principle.²⁹ For better or for worse, justice as fairness is a fully directive, and hence extremely strong, conception of distributive justice.

3. Justice as Fairness as Purely Procedural

3.1 The Rationale for Optimization

That justice as fairness is optimizing and hence fully directive does not as such entail that it is mistaken. To assess its plausibility we need to investigate how and why Rawls endorses a fully directive conception of justice. I begin with the question of how Rawls arrives at such a conception. In my view, the argument from the original position yields a fully directive output because it attributes a fully directive motivation to the parties to the original position: they want to secure as many goods as they can, and as a result they have extremely fine-grained preferences among candidate conceptions of justice. It is not always obvious which of two candidates they prefer, just as it is not always obvious which of two candidate social structures better satisfies the

difference principle or the principle of utility. But we know they will have a preference between any two candidates – again excepting ties, which will be rare – by virtue of the motivations we attribute to them. The argument from the original position is thus a case of optimization in, optimization out. The question is: what justifies framing the argument from the original position in this way?

According to Rawls, the parties to the original position desire as many primary goods as possible. These are all-purpose means that enhance a person's prospects regardless of her plan of life. They are, in Rawls's apt phrase, "what a person wants, whatever else he wants".³⁰ Since we need not appeal to any contingent facts about a person to know that she always has reason to want more primary goods, attributing a desire for more primary goods to the parties to the original position appears to be part and parcel of attributing rationality to them.³¹

To see how this may help defend Rawls's position, consider his example of fair division of a cake between two persons.³² The procedure according to which one person cuts and the other chooses is a fair one, since it produces the independently specifiable fair outcome, where each person receives half the cake. One way to understand what makes this the fair outcome is to attribute to each person a desire for as much cake as possible and then to ask each, subject to the information constraints of the original position, to choose a procedure for dividing the cake. In this case a just principle of division is fully directive, even though it is a principle of justice in the first virtue sense: the parties to this original position would choose the principle that each receive exactly half the cake, and then would choose a procedure such as one-cuts-the-other-chooses in order to satisfy that principle.

3.2 Pure Procedural Justice

It is not clear that Rawls can avail himself of this analogy, however, since on his view societal distributive justice differs crucially from cake division. On Rawls's view a just society's distribution of primary goods is a matter of *pure procedural justice*: so long as the society's

institutions are in fact regulated by the correct conception of justice, the distribution of primary goods they produce is just. Just institutions confer justice on the outcomes they produce, not the other way around.³³ As Rawls observes, cake division is not like this. If an earthquake hits while the first person is cutting the cake, and as a result a cut is made that clearly fails to divide the cake evenly, then it is more just to redivide the cake rather than accept the outcome of the procedure.³⁴ If an earthquake at sea causes the fishing industry to suffer in a way that affects the distribution of social primary goods by the basic structure, by contrast, Rawls would claim that the just outcome is that which the basic structure actually produced, and not that which the basic structure would have produced had it been designed in anticipation of the earthquake.³⁵

This is because justice as fairness is, as Rawls puts it, an *ideal social process* theory.³⁶ Such a theory focuses on “the regulations required to maintain background justice over time”, so that freely chosen social transactions are not infected with injustice by differences in the social position of those transacting.³⁷ Rawls does not think that persons are entitled, full stop, to the share of the social product that the difference principle prescribes. His view is, rather, that labor agreements, contracts more generally, elections, and other important social transactions are fully just only when they take place as part of a social process that is structured so as to satisfy the difference principle. Background justice consists in having a basic structure is organized in this way, and is not vitiated by unforeseen events that occur in the actual process of production and distribution. He believes we should design the basic structure in anticipation of foreseeable accidents, of course, but his criterion of distributive justice is the social background of each transaction and not the distribution in fact produced by the aggregation of all transactions.

This approach contrasts with what Robert Nozick calls “end-state principles”, which focus on the distribution of goods at each moment, without regard to the historical processes that produced this distribution.³⁸ For Rawls it makes all the difference if a person’s lacking goods is a consequence of his having gambled them away, for example, as opposed to his having severely restricted employment opportunities. The latter reflects the distribution of goods by the social

process within which he labors, while the former reflects only his decision about how to use the goods to which he is entitled.

Ideal social process theories also contrast with *ideal historical process* views, such as Nozick's own.³⁹ According to the ideal historical process approach, a just initial distribution conjoined with any number of locally just transactions among individuals always yields a just later distribution. Rawls dissents from this view as well, writing:⁴⁰

Even though the initial state may have been just, and subsequent social conditions may have been just for some time, the accumulated results of many separate and seemingly fair agreements entered into by individuals and associations are likely over an extended period to undermine the background conditions required for free and fair agreements.

I agree with Rawls that this is the best way to frame a theory of justice. I now attempt to use this way of framing Rawls's subject matter to illuminate the central subject matter of this essay: the contrast between the first virtue interpretation of justice as fairness and the competing realistic utopia interpretation.

The realistic utopia understanding of distributive justice, understood as a search for an ideal social process theory, asks for an account of when background social conditions are as fair as they can be. Understood this way, it is not difficult to see how to motivate Rawls's second principle of justice. Even very minor differences in opportunities or in the distribution of goods could affect the justice of procedures like elections and labor markets. If one person has access to a school another cannot afford, or simply has more money than someone else, this could give her an advantage in these competitive arenas, even if they are otherwise fair.

The first virtue understanding of the subject matter of justice, by contrast, asks for an account of when background social conditions are *fair enough* that all individuals can interact on terms of full mutual respect. It is not credible to claim that, to satisfy this norm, background conditions must be optimally fair. A very minor difference in opportunities or goods need not make impossible a relationship of full mutual respect.

This does not mean that there is no way to motivate the difference principle with the normative force of the first virtue. One could withdraw Rawls's commitment to distributive justice as purely procedural, and instead regard distributive justice as an instance of *imperfect procedural justice*.⁴¹ In imperfect procedural justice, it is possible to know in advance what a just or unjust outcome consists in, but it is not possible to design a procedure that will always produce just outcomes. Rawls's paradigm of this phenomenon is a criminal trial. A criminal trial is just if the guilty are found guilty and the innocent are found not guilty, and the aim of producing these results regulates the design of criminal procedures; but there is no set of procedures which always yields the desired results. Instances of imperfect procedural justice are appropriately regulated by end-state principles.

If we understand distributive justice as a case of imperfect procedural justice, then we can avail ourselves of the analogy with the case of cake division. We could think of the entire social product as cooperatively produced in such a way that each member of society has a pro tanto claim to an exactly equal share of this product, regardless of whether this share is distributed by the basic structure of society or produced by individual actions undertaken against the backdrop of the basic structure. This provides a rationale for a fully directive principle of distribution with the normative force of the first virtue, and if we affirm Pareto optimality it provides a rationale for the difference principle in particular.

But this strategy abandons the conception of distributive justice as necessary to ensure background justice for social interactions, including especially elections and labor markets, that is so central to Rawls's project. The difficulty is to see how Rawls can maintain, as he appears to want to maintain, that distributive justice is *both* purely procedural *and* fully directive.⁴² Since the difference principle is fully directive, we can have an extremely specific understanding of the distribution of goods needed to satisfy it.⁴³ This makes it odd to claim that a just distribution is the one that actually obtains, when an accident in the functioning of an otherwise just set of social institutions fails to produce the distribution that the designers of the basic structure would have

produced, had they anticipated the accident. This is what we say about cake division: in response to the earthquake, the person who chooses should either give up some of the cake or offer to redivide it. The fully directive nature of the principle of distribution in this example enables individuals to correct for any accidental errors in the procedure they carry out. Rawls denies, however, that individuals have such extra-institutional obligations to correct for accidental errors in the procedures they carry out to achieve societal distributive justice. There would be no puzzle here, if justice as fairness were not fully directive, and hence were much less specific in content; it would then be unclear what individuals could do that would count as acting unilaterally to better achieve distributive justice.⁴⁴ But as we have seen, Rawls's view is fully directive, so there is a puzzle for his view.⁴⁵

To summarize, then, the tension I see in justice as fairness is as follows. If procedures' typical outcomes determine their justice, as Rawls asserts, then it is natural to think that justice consists in the obtaining of these outcomes, however they are produced; but this contradicts Rawls's understanding of distributive justice as purely procedural.⁴⁶ And conversely, if the justice of outcomes is determined by the justice of the procedures that produce them, as Rawls also asserts, then it is natural to think that a range of outcomes can be just; but this contradicts the fully directive content of justice as fairness. I suggest we resolve this tension by distinguishing between requirements of justice, which constitute the first virtue understanding of justice, and ideals of justice, which constitute the realistic utopia understanding. With this distinction in view a new Rawlsian position can be articulated. On this position, satisfaction of the second principle of justice is *not* required in order for transactions among individuals to be adequately fair, and so procedures can confer justice on their results even if they fail to satisfy the difference principle. Complete satisfaction of the second principle remains a plausible candidate for optimally just background conditions, however, and so justice as fairness in its entirety can be asserted as a realistic utopia.⁴⁷

4. The Argument from the Original Position Reconsidered

4.1 The Argument Against the Principle of Average Utility

In Parts 2 and 3 I argued that it is problematic to advance a fully directive conception of justice with the normative force of the first virtue of social institutions, and I argued that doing so stands in tension with Rawls's understanding of distributive justice as a matter of pure procedural justice. I suggested that the best Rawlsian response to this tension is to withdraw a commitment to the claim that justice as fairness has this normative force. A Rawlsian can preserve the appeal of his theory while asserting only constitutional essentials with this force, asserting the second principle of justice instead merely as a realistic utopia.

I now turn to consider how Rawls's arguments for justice as fairness fare under this new understanding of the theory. I focus especially on the arguments designed to vindicate the second principle of justice.⁴⁸ Rawls's case for justice as fairness is the famous argument from the original position. In Part 2 I described the information constraints faced by the parties to the original position and the motivations that are attributed to them. Although the parties face these restrictions on available information, they know all general and uncontroversial facts of human social science, including chiefly that humans are in what David Hume called the "circumstances of justice".⁴⁹ The parties reach a decision about their preferred conception of justice through a series of pairwise comparisons: they are offered two conceptions of justice, and they indicate which they prefer.

Rawls focuses his attention on two such comparisons: that between justice as fairness and the *principle of average utility*, and that between justice as fairness and the *principle of restricted utility*.⁵⁰ The principle of average utility states that socially produced goods should be distributed so as to maximize the total happiness in a society divided by the number of individuals in that society. The principle of restricted utility also asserts a version of the principle of average utility, but it restricts this principle by lexically subordinating it to a prior principle which states that adequate constitutional essentials must be realized; the principle of restricted utility thus states

that socially produced goods should be distributed so as to maximize average happiness subject to the prior constraint that adequate constitutional essentials are realized.⁵¹

The crux of the argument that the parties to the original position would prefer justice as fairness over the principle of average utility is that the latter principle fails to guarantee citizens the constitutional essentials: it is a contingent matter, even under favorable historical conditions, whether satisfaction of the principle of average utility provides these constitutional essentials to everyone. There are circumstances where the principle of average utility permits an unequal distribution of political liberties, and there are circumstances where this principle permits a distribution of opportunities or material goods inadequate to meet citizens' urgent needs. Citizens who are apportioned fewer political liberties than their fellow citizens, who have inadequate political and economic opportunities to take advantage of, or who live in poverty will experience what Rawls calls great "strains of commitment" as they attempt to meet the demands of justice so understood.⁵² According to Rawls, these strains of commitment typically engender bitterness in humans forced to undergo them. This bitterness in turn makes citizens less apt to live up to the demands of justice and more prone to radical or revolutionary political positions.⁵³ The parties to the original position fear both undergoing these strains of commitment and the political instability that characteristically results when their fellow citizens undergo these strains.

This is a powerful, and I think decisive, argument against the principle of average utility as a conception of justice. Notice that nothing in this argument trades on the details of typical utilitarian views. The argument trades neither on utilitarianism's consequentialism, nor on its optimization requirement, nor on its characteristic appeals to desire- or preference-satisfaction as a currency of justice. All the argument needs as a premise is that the principle of average utility fails to guarantee the constitutional essentials. For this reason, the argument can be extended to explain why the parties would select justice as fairness not only over average utilitarianism but also over any other rival conception that fails to guarantee the constitutional essentials.

4.2 The Argument Against the Principle of Restricted Utility

The second principal argument from the original position asks the parties to compare justice as fairness with the principle of restricted utility, which is an instance of what in *A Theory of Justice* he calls a “mixed” conception of justice.⁵⁴ The principle of restricted utility states that the principle of average utility should be satisfied, but only subject to the restriction that the constitutional essentials are realized. The principle of restricted utility is more plausible than the principle of average utility as a conception of justice, in part because it is designed to avoid the first argument from the original position. Satisfying the principle of restricted utility involves satisfying the equal liberties principle and securing adequate opportunities and material goods for all citizens. As Rawls acknowledges, this makes the parties’ choice between it and justice as fairness a much closer decision.⁵⁵ At stake are the relative merits of the principle of average utility and Rawls’s second principle, considered not as complete understandings of distributive justice, but only as principles for distributing social goods other than constitutional essentials.

Rawls advances three families of considerations in favor of justice as fairness over the principle of restricted utility: considerations of publicity, of reciprocity, and of stability. Publicity obtains to the extent that the conception of justice regulating the basic structure is known by all, to the extent to which the basic structure is known to meet the demands of that conception of justice, and to the extent that the best justification for that conception of justice is available in the society’s public political culture.⁵⁶ Reciprocity obtains to the extent that the conception of justice regulating the basic structure works to the mutual advantage of all citizens. Stability obtains to the extent that the conception of justice regulating the basic structure would be recognized as good by humans living under and with that conception of justice; and so would come to adopt it as their own conception of justice and thus freely perpetuate it over time.

These three families of considerations are progressively more demanding.⁵⁷ If publicity fails to obtain to a high degree, this threatens reciprocity; not only must the basic structure work to the mutual advantage of all citizens, it must be seen to do so, if it is to be robustly reciprocal

and hence robustly cooperative. There are further conditions on reciprocity, however, apart from publicity. A conception of justice where some political liberties are accorded to one social class but not to another may be perfectly public, but such a conception fails to work to the mutual advantage of members of both classes. Whenever reciprocity fails to obtain to a high degree, this threatens stability; for Rawls believes humans typically develop and exercise a sense of justice only to the extent that their society that is robustly cooperative.⁵⁸ But stability requires not only that the basic structure of a society be effectively regulated by a public and reciprocal conception of justice, it requires also that the humans who live in that society be able to appreciate what is good about the prevailing conception of justice and so come to adopt it as their own. It is an open question of human psychology whether all conceptions of justice that are robustly reciprocal will be such as to garner allegiance from humans who live under and with them. I will not distinguish here between Rawls's arguments from reciprocity and those from stability, since the latter extend beyond the former by taking up controversies in the theory of human psychological development that are outside the scope of this essay.

According to Rawls, two facts are insufficiently public under the principle of restricted utility: the distribution of opportunities and material goods that constitutes realization of the constitutional essentials, and the distribution of social goods beyond the constitutional essentials that would optimize average utility.⁵⁹ The latter fact is insufficiently public because the currency of the principle of restricted utility is desire- or preference-fulfillment; the extent of desire- or preference-fulfillment is difficult to measure, and the question of what trade-offs of goods among individuals are permitted is a notorious problem for utilitarianism. Under justice as fairness, by contrast, the currency of distributive justice is social primary goods, which are by design more public and easier to measure. Under justice as fairness, furthermore, it is not necessary to identify the distribution of opportunities and material goods that minimally realizes the constitutional essentials. The second principle of justice demands that these goods be distributed to the least-advantaged group well above this threshold, so satisfaction of Rawls's second principle entails

that the constitutional essentials have been provided to all citizens, regardless of where exactly the relevant threshold lies.⁶⁰

Note that these arguments appear to presuppose the first virtue understanding of the subject matter of justice. If the dispute between justice as fairness and the principle of restricted utility concerns the conditions under which a person is violated or denigrated, it is indeed of crucial importance to identify the precise conditions under which the principle of restricted utility is satisfied. It is also important to identify the precise conditions under which the principle of average utility kicks in, and so we must also identify precisely the minimum threshold where the constitutional essentials have been provided. I have already argued against the view that any fully directive conception of distributive justice has the normative force of moral obligation, however, so it is worth asking what force Rawls's arguments have when they are construed as arguments only for justice as fairness as a superior regulative ideal or realistic utopia.

When so understood, the argument from publicity turns out to be quite weak. On this understanding, the advocate of the principle of restricted utility now claims only that satisfaction of that principle is the most just social arrangement, not more strongly that any other arrangement is impermissible or intolerable. She can therefore address the difficulty of locating the point at which the constitutional essentials are realized by allowing that we should err on the side of the least-advantaged group when making this difficult judgment. This may come at some cost to average utility. But since providing the constitutional essentials to all citizens has the normative force of obligation, while average utility has the normative force only of a regulative ideal, she has reason to assert that the former must be clearly satisfied before we seek to satisfy the latter. Similarly, the difficulty of identifying the point at which the principle of restricted utility is satisfied is much less problematic on the realistic utopia construal of the principle of restricted utility. Less is riding on our success in identifying this point, since no one is as such violated, denigrated, or otherwise wronged by a failure to identify it correctly. Approximate satisfaction of the principle may be perfectly adequate.

The argument from reciprocity fares better, however, under the realistic utopia construal of the subject matter of justice. The fair equality of opportunity principle and the difference principle are indeed more robustly cooperative than the principle of restricted utility, since the latter permits the least-advantaged group to be made worse off for the sake of compensating benefits for more advantaged members of society. A society regulated by justice as fairness embodies a more mutually respectful and mutually accountable citizenry than one regulated by the principle of restricted utility.

Understanding justice as fairness only as a realistic utopia qualifies the conclusion, of course, of the argument from reciprocity. It does not follow from this conclusion, as it would from the conclusion that justice as fairness is the first virtue of social institutions, that we should always try to better realize this conception. The argument, even if successful, establishes only that a society is more just to the extent that it better realizes this conception.⁶¹ Competing considerations outside the domain of justice may justify departures from the greater justice than would obtain under justice as fairness. Perhaps there is justice-independent value in rewarding productive activity. Perhaps there is educational value – or even intrinsic value – for humans in having contexts where they have agonistic relationships with their fellow citizens. Or perhaps there is value for humans in the development of talents or the achievement of excellence that justifies departures from satisfaction of the difference principle. I will not here attempt to adjudicate these issues, but I repeat that they are not ruled out by the new interpretation of the argument from reciprocity, even if this argument establishes that justice as fairness is superior to the principle of restricted utility as a conception of justice.⁶² I am not trying here to show that the interpretation of justice as fairness I offer is our best available regulative ideal; I am attempting to show only that it dissolves a tension in Rawls's views and that it is independently attractive.

I mentioned earlier that bitterness is the characteristic human emotional response to a failure to provide constitutional essentials, according to Rawls, and that attraction to radical or revolutionary political movements is a typical manifestation of this bitterness. By contrast, Rawls

claims that the characteristic human emotional response to a failure to achieve robustly reciprocal relationships among citizens is alienation, and he claims further that disillusionment with the political process and apathetic failure to participate in it are the typical manifestations of this alienation.⁶³ The test for whether justice as fairness is our realistic utopia is whether all of its competitors, and not only the principle of restricted utility, encourage this alienation to a greater degree. Even if this is true – I am not sure whether it is – this does not suffice to establish, on the interpretation of justice as fairness I offer here, that we should aim to reform our society in the direction of justice as fairness. For even if all other sets of regulative principles are less just, and engender more alienation, we must establish in addition that this opportunity cost of justice and alienation is not worth the price of the achievement of other values.

5. Conclusion

In this essay I have explored a tension between two different ways that John Rawls characterizes a conception of distributive justice: as the first virtue of social institutions, where principles of justice have the normative force of moral prohibitions, and as a realistic utopia, where principles of justice have the normative force of regulative ideals. I have argued that it is problematic to advance a fully directive conception of justice with the normative force of the first virtue of social institutions, and that doing so stands in tension with Rawls's understanding of distributive justice as a matter of pure procedural justice. I have suggested that the best Rawlsian response to this tension is to withdraw the claim that justice as fairness taken as a whole has this normative force. A view can preserve the appeal of Rawls's theory by asserting constitutional essentials alone with this normative force and asserting the second principle of justice only as a realistic utopia.

Works Cited

- Cohen, G. A. (1997). "Where the Action Is: On the Site of Distributive Justice". *Philosophy & Public Affairs* 26, 1997.
- _____ (2008). *Rescuing Justice and Equality*. Harvard, 2008.
- Cohen, Joshua (2001). "Taking People as They Are". *Philosophy & Public Affairs* 30, 2001.
- Daniels, Norman (1975). "Equal Liberty and Unequal Worth of Liberty". In *Reading Rawls*. Ed. Norman Daniels. Stanford, 1975.
- Darwall, Stephen (1976). "A Defense of the Kantian Interpretation". *Ethics* 86, 1976.
- Estlund, David (1998). "Liberalism, Equality, and Fraternity in Cohen's Critique of Rawls". *Journal of Political Philosophy* 6, 1998.
- Frankfurt, Harry (1987). "Equality as a Moral Ideal". *Ethics* 98, 1987.
- Freeman, Samuel (1994). "Utilitarianism, Deontology, and the Priority of Right". *Philosophy & Public Affairs* 23, 1994.
- _____ (2007a). *Rawls*. Routledge, 2007.
- _____ (2007b). "Consequentialism, Publicity, Stability, and Property-Owning Democracy". In *Justice and the Social Contract*. Oxford, 2007.
- Habermas, Jürgen (1992). *Between Facts and Norms*. Trans. William Rehg. MIT, 1996.
- _____ (1995). "Reconciliation Through the Public Use of Reason". Trans. Ciarin Cronin. *Journal of Philosophy* 92, 1995.
- Harsanyi, John (1953). "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking". *Journal of Political Economy* 61, 1953.
- Hart, H. L. A. (1973). "Rawls on Liberty and Its Priority". In *Reading Rawls*. Ed. Norman Daniels. Stanford, 1975.
- Hume, David (1739). *A Treatise of Human Nature*. Penguin Books, 1969.
- Johnson, Oliver (1974). "The Kantian Interpretation". *Ethics* 84, 1974.
- Julius, A. J. (2003). "Basic Structure and the Value of Equality". *Philosophy & Public Affairs* 31, 2003.
- Kant, Immanuel (1785). *Groundwork of the Metaphysics of Morals*. Trans. and Ed. Mary Gregor. Cambridge, 1997.
- Locke, John (1690). *Second Treatise of Government*. Ed. C. B. Macpherson. Hackett, 1980.
- Murphy, Liam (1998). "Institutions and the Demands of Justice". *Philosophy & Public Affairs* 27, 1998.
- Musgrave, R. A. (1959). *The Theory of Public Finance*. McGraw-Hill, 1959.
- Nagel, Thomas (1986). *The View from Nowhere*. Oxford, 1986.
- _____ (1991). *Equality and Partiality*. Oxford, 1991.
- Nozick, Robert (1974). *Anarchy, State, and Utopia*. Basic Books, 1974.
- Pogge, Thomas (1995). "Three Problems with Contractarian-Consequentialist Ways of Assessing Social Institutions". *Social Philosophy & Policy* 12, 1995.
- _____ (2000). "On the Site of Distributive Justice: Reflections on Cohen and Murphy". *Philosophy & Public Affairs* 29, 2000.
- _____ (2007). *John Rawls*. Trans. Michelle Koch. Oxford, 2007.
- Rawls, John (1971). *A Theory of Justice*. Harvard, 1971.
- _____ (1982). "Social Unity and Primary Goods". In *Collected Papers*. Ed. Samuel Freeman. Harvard, 1999.
- _____ (1993). *Political Liberalism*. Columbia, 1993.
- _____ (1995). "Reply to Habermas". In *Political Liberalism*, Second Edition. Harvard, 1996.
- _____ (2001). *Justice as Fairness: A Restatement*. Ed. Erin Kelly. Harvard, 2001.

- Scanlon, T. M. (1996). "The Diversity of Objections to Inequality". In *The Difficulty of Tolerance*. Cambridge, 2003.
- Scheffler, Samuel (1982). *The Rejection of Consequentialism*. Oxford, 1982.
- Shiffrin, Seana (1991). "Moral Autonomy and Agent-Centred Options". *Analysis* 51, 1991.
- Slote, Michael (1985). *Common-Sense Morality and Consequentialism*. Routledge, 1985.
- Van Parijs, Philippe (2003). "Difference Principles". In *The Cambridge Companion to Rawls*. Ed. Samuel Freeman. Cambridge, 2003.
- Waldron, Jeremy (1986). "John Rawls and the Social Minimum". *Journal of Applied Philosophy* 3, 1986.
- Williams, Andrew (1998). "Incentives, Inequality, and Publicity". *Philosophy & Public Affairs* 27, 1998.
- Williams, Bernard (1973). "A Critique of Utilitarianism". In J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against*. Cambridge, 1973.

¹ Rawls (1971), 3.

² Rawls (1971), 253.

³ Rawls (2001), 4-5.

⁴ Rawls (1971), 8-9, and Rawls (2001), 13, 65-66. The term “ideal theory” is sometimes used to refer to the theory of ideals and sometimes used to refer to the theory of conditions where there are no violations of justice in the first virtue sense. To avoid confusion between these two different uses, I recommend using the term only in the former way, and instead using “satisfactory theory” for the latter concept. For further development of the distinction between ideal and satisfactory theory, see [citation omitted for anonymity].

⁵ This distinction arises in normative domains other than justice. There are “first virtue” norms of honesty, for example, whose violation entails impermissibility, and there are also regulative ideals of honesty whose violation lacks this entailment.

⁶ Rawls (1993), 227-230, and Rawls (2001), 47-50.

⁷ Rawls famously labels the currency of justice “social primary goods”. For Rawls’s account of these goods, see Rawls (1971), 90-95, Rawls (1982), Rawls (1993), 178-190, and Rawls (2001), 57-61.

⁸ For detailed expositions of Rawls’s work, see Samuel Freeman (2007a) and Thomas Pogge (2007). For specific discussion of the role of the domain of right and wrong in justice as fairness, see Freeman (1994).

⁹ Rawls (2001), 42-43.

¹⁰ Rawls states explicitly that the difference principle is an optimizing principle; see Rawls (1971), 79.

¹¹ See Shiffrin (1991).

¹² See, for example, Bernard Williams (1973), Samuel Scheffler (1982), Chapters 1 and 2 of Michael Slote (1985), Chapter IX of Thomas Nagel (1986), and Shiffrin (1991).

¹³ Rawls is sensitive to some of the dangers of advancing fully directive normative principles. By Rawls’s own account, the most significant change to the content of justice as fairness over time is that the equal liberties principle is amended from an optimizing formulation to a non-optimizing formulation. See Rawls (1993), 289-294; and for criticism of the optimizing formulation see H. L. A. Hart (1973) and Norman Daniels (1975).

¹⁴ For a full development of this criticism, see [citation omitted for anonymity]. For Rawls’s comments on discretionary government spending, see Rawls (1971), 282-284 and 331-332.

¹⁵ T. M. Scanlon has an extremely helpful discussion of the reasons for advancing an egalitarian conception of distributive justice, which bears on whether egalitarianism is most plausible when it is formulated as a fully directive principles; see Scanlon (1996). Scanlon’s discussion is in part a reaction to Harry Frankfurt (1987). For Rawls’s case for strong egalitarianism, see Rawls (2001), 130-132.

¹⁶ See Rawls’s discussion of what he calls “branches of government” at Rawls (1971), 274-284. He adopts this term from R. A. Musgrave (1959).

¹⁷ For Rawls’s account of the basic structure, see Rawls (1971), 7-11, Lecture VII of Rawls (1993), and Rawls (2001), especially 52-57.

¹⁸ Rawls (1971), 275-280.

¹⁹ Note that the division of labor among institutions of the basic structure I discuss here is not the same as the division of labor between the basic structure and individuals endorsed by Rawls and others, including Nagel; see Chapter 6 of Nagel (1991). For a criticism of this other division of labor, see Liam Murphy (1998), and for defenses of it see A. J. Julius (2003) and [citation omitted for anonymity].

²⁰ To make this claim vivid, one might consider, by way of comparison, an individual who aims to satisfy only approximately a principle of individual action that he acknowledges both to be valid and to have the normative force of a Kantian categorical imperative.

²¹ It is perhaps useful to compare Rawls’s lack of emphasis on the precise conditions under which the difference principle is satisfied with the lack of emphasis utilitarians sometimes exhibit towards the precise conditions under which the principle of utility is satisfied. The latter attitude may be justified by the fact that the utilitarian endorses a deflated conception of right and wrong. This avenue appears unavailable to Rawls, however, since he resists such deflationism about the right. This lack of emphasis on the precise conditions of satisfaction of justice as fairness is an important strand in Rawls’s thought away from the first virtue understanding of justice and towards the realistic utopia understanding.

²² [citation omitted for anonymity]

²³ Rawls famously labels this exclusion of information the “veil of ignorance”. For his account of the veil of ignorance, see Rawls (1971), 12, 136-142, and Rawls (2001), 15, 18.

²⁴ For Rawls's account of the four-stage sequence, see Rawls (1971), 195-201, Rawls (1995), and Rawls (2001), 48.

²⁵ See Rawls (1971), 199-201, and Rawls (2001), 48. It is not clear from Rawls's exposition of the four-stage sequence whether any additional information, apart from knowledge of the constitution, is available at the third stage but not available at the second.

²⁶ According to Rawls, a commitment to some principle of equal opportunity and to the meeting of urgent humanitarian needs may be included in the constitution, even though these commitments pertain to the distribution of goods other than liberties. But he claims that nothing as specific as the fair equality of opportunity principle or the difference principle should be included in the constitution. See Rawls (1993), 228-229, and Rawls (2001), 162.

²⁷ Rawls is explicit that the choice between public and private ownership of the means of production is not settled by justice as fairness alone, and this may be thought to support the claim that there is a wide range of options available at the second stage of the sequence, before the difference principle is ever applied; see Rawls (1971), 270-274, and Rawls (2001), 136-140. But this is a misreading of Rawls. While it is true that in the abstract justice as fairness fails to settle this question of large-scale economic structure, the theory *does* settle this question when conjoined with a society's historical circumstances, such as its culture and available resources and technologies. In any particular context of application, then, justice as fairness will underdetermine an answer to this question only in the highly unlikely event that each system would, in that context, yield the exact same level of distribution to the least-advantaged group. For further development of this point, see [citation omitted for anonymity].

²⁸ There must be such a constitution, if there is latitude at the second stage of the sequence; and there will be such latitude, since the equal liberties principle is not an optimizing principle. For references on this point, see note 13 above.

²⁹ For more on this point, see [citation omitted for anonymity]. Note that Rawls's common characterization of the constitution as the primary vehicle of societal cooperation does not undermine the argument in the text, since the ability to amend the constitution is part of any plausible understanding of such cooperation. Note also that the four-stage sequence is a heuristic that anyone can engage in at any time, so there is no principled bar to moving back and forth through the stages; see Rawls (1971), 200-201. And finally, it is worth noting the similarity of this line of reasoning with Rawls's own reasoning when he articulates the most complete statement of the difference principle at Rawls (1971), 82-83. G. A. Cohen calls this the "canonical" version of the difference principle; he criticizes both it and the shortened version I work with here in Chapter 4 of Cohen (2008). For another important recent discussion of the difference principle, see Philippe van Parijs (2003).

³⁰ Rawls (1971), 92.

³¹ One might object that Rawls is mistaken about this claim by comparing it to the analogous claim in the individual case. According to Immanuel Kant, for example, every person desires happiness, regardless of her plan of life; but of course it does not follow that the pursuit of happiness, subject to the constraints of morality, itself has the normative force of a categorical imperative. See Kant (1785), 26-29 or 4:415-419. It could be, however, that justice differs from individual moral requirements on this score in a way that vindicates Rawls. For a related criticism of Rawls, see Oliver Johnson (1974); and for a defense of Rawls against this criticism, see Stephen Darwall (1976).

³² See Rawls (1971), 85, and Rawls (1995), 421-423.

³³ Rawls (1971), 86-89. G. A. Cohen's criticisms of Rawls are appropriately celebrated, but in my view he overstates the extent to which these criticisms are internal to Rawls's theory, since I believe that Rawls's commitment to distributive justice as purely procedural is deeper than his commitment to the substantive content of the difference principle. For the most recent incarnation of Cohen's critiques, see Cohen (2008).

³⁴ For some concerns about this feature of Rawls's theory, see Pogge (1995).

³⁵ Rawls (2001), 50-52.

³⁶ See Rawls (2001), 52-55.

³⁷ Rawls (2001), 54.

³⁸ See Nozick (1974), 153-155.

³⁹ Nozick (1974). Both Nozick and Rawls attribute this approach to John Locke; see Chapter V of Locke (1690).

⁴⁰ Rawls (2001), 53.

⁴¹ Rawls (1971), 85-86.

⁴² For a related criticism of Rawls, see Cohen (2008), 156-161. As I stated in earlier footnote, I believe Cohen is mistaken to claim that Rawls's commitment to the fully directive and strongly egalitarian content of the difference principle is deeper than his commitment to distributive justice as the purely procedural securing of background justice for social interactions like elections and labor markets. This does not refute Cohen's arguments, of course, it simply situates them differently with respect to Rawls. No doubt there are many Rawlsians who hold the views Cohen attributes to Rawls, moreover, and they are susceptible to his arguments as he presents them.

⁴³ Nozick famously criticizes this feature of Rawls's view; see Chapter 7 of Nozick (1974).

⁴⁴ If there is a failure to realize even the constitutional essentials, then it is clear that individuals must do what can be done to address this failure. But this is best understood as a humanitarian obligation, rather than as rectification of failure of distributive injustice.

⁴⁵ Rawls's famous reply to Jürgen Habermas may be thought to encourage a reading where he abandons his commitment to distributive justice as purely procedural. Rawls writes that "the justice of a procedure always depends (leaving aside the special case of gambling) on the justice of its likely outcome"; see Rawls (1995), 421, and the parentheses are Rawls's. This may appear to be an admission that, with the possible exception of lotteries and the like, justice is never purely procedural, but rather always consists in an effort to realize an outcome that is independently specifiable as just. But this is a misreading of the passage in question. Given how often and how strongly Rawls asserts his conception of distributive justice as purely procedural, including especially the fact that he does not withdraw this assertion in *Justice as Fairness: A Restatement*, it cannot be that this passage from Rawls's reply to Habermas indicates a major shift in his position. It would be very strange, moreover, for Rawls to intend this passage to set out a new direction when he makes no explicit indication of this and when the tone of the reply is clearly one of defense of his previously articulated positions. We should instead read Rawls as claiming only that we regulate the basic structure in part by the outcomes it tends to produce. He does not intend this to imply that the outcomes we seek to produce are just independent of their having been produced by a basic structure. To produce these outcomes in other ways, for example by direct individual action, is not, on Rawls's view, to realize justice. Rawls's conception of distributive justice is not, by contrast to its common characterization by opponents, focused on end-states. End-states are just only insofar as they are produced by a system with the principal moral function of realizing and preserving background justice so that consensual social relationships among individuals will be just. For Habermas's critique of Rawls, see Habermas (1995), and for Habermas's own procedural account of justice, see Habermas (1992).

⁴⁶ This theme is developed at length in Cohen (2008).

⁴⁷ I repeat here that I do not claim this to be Rawls's own view; there are passages where it appears clear that he understands the second principle to have the normative force of the first virtue. See, for example, Rawls (2001), 129. But in view of the tension I discuss, I think Rawlsians should retract this claim.

⁴⁸ See Section 29 of Rawls (1971) and Part III of Rawls (2001).

⁴⁹ The circumstances of justice include the condition of "moderate scarcity", where goods are neither so abundant that all desires can be satisfied nor so scarce that they are exhausted meeting humanitarian needs. They also include the condition of "mutual disinterest", where people are understood to be capable of acting for the sake of others' interests but not typically motivated to do so unless they stand in a special relationship to these others. See Hume (1739), 545-547, Rawls (1971), 126-130, and Rawls (2001), 84-85.

⁵⁰ These rival conceptions of justice are in fact families of rival conceptions, since different utilitarian views have importantly distinct understandings of what utility consists in; for advocacy of the principle of average utility, see John Harsanyi (1953). The principle of total utility is dismissed out of hand by Rawls, and he believes that the argument from the original position reveals it to have a wholly different, and far less plausible, rationale than that for the principle of average utility. See Rawls (1971), 161-166.

⁵¹ In *Justice as Fairness: A Restatement*, Rawls says the principle of restricted utility provides a "suitable social minimum", rather than saying this principle guarantees constitutional essentials. See Rawls (2001), 120. But this is not different from what I assert in the text, for as Rawls uses these terms "some principle of opportunity is a constitutional essential ... [and] a social minimum providing for the basic needs of all citizens is also a constitutional essential". See Rawls (2001), 47-48. Note that in *A Theory of Justice* Rawls uses "social minimum" in a different way, to pick out a level of provision (typically well beyond what is required as a constitutional essential) that is guaranteed under the difference principle. See Rawls (1971), 276-277, 285-286, 304; and see also Jeremy Waldron (1986), for an excellent critical discussion of Rawls's understanding of the social minimum.

⁵² See Rawls (1971), 176-178, and Rawls (2001), 128-130.

⁵³ Rawls (2001), 128.

⁵⁴ See Rawls (1971), 315-318, and Rawls (2001), 119-130

⁵⁵ Rawls (2001), 95-96.

⁵⁶ See Rawls (1993), 66-71, and Rawls (2001), 120-121. For discussion of the publicity requirement, see Freeman (2007b). See also Andrew Williams (1998); Williams defends Rawls against critiques due to Cohen (1997). I do not have space to discuss this critique explicitly here, but for further discussion see David Estlund (1998), Murphy (1998), Pogge (2000), Joshua Cohen (2001), Julius (2003), and [*citation omitted for anonymity*].

⁵⁷ To the best of my knowledge Rawls does not state this explicitly, but I believe it is his view. The best competing interpretation would have it that publicity and reciprocity are mutually independent, but that each is a constituent (though not jointly exhaustive) of stability. For the reasons indicated in the text, I believe against this that all considerations of publicity, as Rawls understands it, are also considerations of reciprocity.

⁵⁸ Rawls (1971), 499f.

⁵⁹ Rawls (2001), 126-129.

⁶⁰ This problem is likely less severe than the other, since it is plausible that needs are, generally speaking, more public than wants. But I do not pursue that hypothesis here, since on my recommended interpretation the argument from reciprocity constitutes the bulk of the case against the principle of restricted utility.

⁶¹ Unlike the argument against the principle of average utility, furthermore, it is less clear that the argument from reciprocity has force against other competing principles. Any conception of justice that fails to realize the constitutional essentials will, for that reason, generate intolerable strains of commitment for citizens living under that conception. But there will be competitors to the second principle of justice that are not clearly less reciprocal. The principle of restricted utility is prone to the charge of insufficient reciprocity because it permits trade-offs of one person's desire-fulfillment for another's. A rival conception of justice that violates Rawls's second principle of justice in order to enhance other values may be less vulnerable to this charge.

⁶² It is worth saying explicitly that this is because a theory of justice, even when understood as a realistic utopia, does not aim to provide an all-encompassing ideal of societal design. For much more on that point see Part II of Cohen (2008).

⁶³ Rawls (2001), 128.