The Review of Politics 87 (2025), 393–412.
© The Author(s), 2025. Published by Cambridge University Press on behalf of University of Notre Dame. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.
doi:10.1017/S0034670525000051

Reciprocity, Self-interest, and Citizen Virtue in Justice as Fairness

S. A. Lloyd
University of Southern California, Los Angeles, CA, USA

Abstract: Does John Rawls's justice as fairness require virtuous citizens to eschew self-interested motivation in favor of motivation by the aim to promote justice when making decisions about their labor or investment participation in their society's economy? Analyzing Rawls's ideas of advantage and of legitimate ends and tracing the purpose and implementation of his difference principle, the article shows how G. A. Cohen's critique of Rawls's allowance of productive incentives reflects misunderstandings of Rawls's theoretical aims and of the fundamental liberalism of Rawls's well-ordered society. In that society, political civility permits making such decisions by appeal to each citizen's self-interested aims and values.

Introduction

To ascertain the truth or falsity of today's widely expressed opinion that there is less civility in citizens' political discourse and civic interaction than there was prior to the digital age would require detailed historical and empirical information in a carefully designed comparative study. Yet it is easy to understand how features of our digital age might reduce barriers to political incivility and intensify our experience of it. The ability to interact anonymously invites such vicious behaviors as misleading, manipulating, bullying, and trolling; the immediacy of digital communication encourages acting without due consideration; social media's echo-chamber effect tempts piling on and cancelling fellow citizens; information silos lead to oversimplification and polarization. All of these may undermine citizens' mutual respect, sense of equal status, and willingness to participate constructively in deliberative democracy.

But why are we tempted to political incivility in the first place, in any age or social environment? Very often it is because we think it will serve our

S. A. Lloyd is Professor of Philosophy and Law in the School of Philosophy at the University of Southern California, Los Angeles, California 90089-0451 (lloyd@usc.edu).

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

perceived self-interest. By self-interest I do not mean only our selfish interests *in* ourselves—my interest in my health, my wealth, my reputation, my pleasure—but also the interests *of* ourselves in states of the world that realize our values, say, my interest in the continued existence of wild elephants and polar bears, or in women's having a legal right to bodily autonomy, or in alleviating the suffering of civilians caught between combatants. Very often our interest in advancing our own ends and values motivates our use of vicious tactics in the political arena.

This article investigates civility's relation to self-interest, that basic, ubiquitous motivation seemingly involved in most intentional departures from political civility. To make this large topic more manageable, I narrow our focus to the requirements of civility in Rawls's just society—the well-ordered society of justice as fairness—and I narrow the domain of self-interested decisions to those about what work or other economically productive activity we will do. Although Rawls eventually came to the view that there might be several different attractive liberal conceptions of justice, and that this whole family of liberal conceptions might be the only feasible object of an overlapping consensus of reasonable comprehensive doctrines, he nevertheless continued to believe that his own "justice as fairness" was the best such conception. Whereas Rawls did repudiate his former discussion in Part Three of *A Theory of Justice* of the basis of stability in the just society, he did not revise his view of what the most just society is. I ask to what extent it is permissible for Rawls's virtuous citizen to deliberate and to act on the basis of self-interest, specifically when determining what productive contribution s/he will make to their society's economy.

In Section 1 I lay out a challenge to Rawls's theory of justice alleging that its allowing the use of economic incentives to motivate citizens' productive behavior is incompatible with realizing the aim of Rawls's difference principle. Section 2 describes some of the ways one person's labor at a time could increase the economic well-being of other persons at that time, thereby seeming to satisfy the difference principle in what I term its synchronic personal application, and then argues that because the difference principle requires reciprocal benefit, it may very well require providing productive incentives to those who create economic benefits for others. Section 3 considers and rejects a variant of the challenge presented in Section 1 that seeks to discredit the difference principle itself. Section 4 gives content to the ideas of citizens who are well-situated, who are less well-situated, and who are badly situated, arguing that in Rawls's system, the mere fact of less well-situatedness does not automatically command the attention of justice. Section 5 shows that Rawls intended a diachronic positional application of the difference principle rather than a synchronic personal application. Because citizens are in no position to affect the justice of society's background basic structure through their own individual choices, they need not act from an internalized ethos of justice, eschewing self-interest, in order to be virtuous citizens. Section 6 reveals Rawls's aim to be construction of a society that

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

enables free and equal citizens to accept the fact that social and economic inequalities are deeply influenced by contingencies and happenstance. The goal of justice is not to eliminate all differences in citizens' life prospects, nor to redress such inequalities. In the Conclusion I review the defects of the initial challenge to Rawls's theory of justice and explain how it is in fact a great strength of Rawls's theory that it does not require virtuous citizens to subordinate their self-interested pursuit of their own aims, ends, and values to an alpha goal of promoting justice.

Noteworthy scholarly treatments of G. A. Cohen's critique of Rawls's acceptance of some productive incentives provide alternative criticisms grounded in philosophical assumptions very different from those investigated here. Jan Narveson faults Cohen's critique for its incompatibility with libertarian principles of justice.[1] Alan Thomas criticizes Cohen's argument for entailing an unreasonably demanding moral rigorism, arguing that Rawls's principles instead commit him to a system like James Meade's revisionary socialism.[2] Paul Smith argues that because the social arrangements recommended by Rawls's theory are consistent with those recommended by Cohen's, Cohen's critique of Rawls's egalitarianism is misplaced, their disagreement being not principled but merely tactical.[3] In contrast, the present investigation exposes important deficiencies in Cohen's understanding of the purpose and operation of Rawls's difference principle in justice as fairness.

1. An Argument for the Impermissibility of Self-Interested Deliberation

It is helpful to begin by considering a challenge to Rawls's theory of justice made by G. A. Cohen.[4] Cohen's general aim was to show that no society can realize justice unless its members internalize and act on an egalitarian "ethos" rather than on self-interest when making personal choices. This means that

[1]Jan Narveson, "Cohen on Rawls on Incentives and Equality," *Socialist Studies, Etudes socialistes* 8, no. 1 (Winter 2012): 35–56.

[2]Alan Thomas, "Cohen's Critique of Rawls: A Double Counting Objection," *Mind* 120, no. 480 (2011): 1099–1141.

[3]Paul Smith, "Incentives and Justice: G. A. Cohen's Critique of Rawls," *Social Theory and Practice* 24, no. 2 (1998): 205–35.

[4]Cohen delivered "Incentives, Inequality, and Community" as the 1991 Tanner Lecture at Stanford University. Citations are to this lecture as it appears in G. A. Cohen, *Equal Freedom; Selected Tanner Lectures on Human Values*, ed. Stephen Darwall (Ann Arbor: University of Michigan Press, 1995). He develops the ideas in G. A. Cohen, "Where the Action is: On the Site of Distributive Justice," *Philosophy & Public Affairs* 26, no. 1 (Winter 1997): 3–30, and in chapters 8 and 9 of G. A. Cohen, *If you're an Egalitarian, How Come you're So Rich?* (Cambridge, MA: Harvard University Press, 2000).

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

political civility—the virtue of citizens—is not merely to sustain just institutions and conform to their rules, but also requires exhibiting an egalitarian ethos in personal choices. In seeking to show this for Rawls's just society, Cohen focused on the effect of personal decisions about work and productive activity on the implementation of Rawls's difference principle. I reconstruct and consolidate Cohen's several sub-arguments into a unified whole which we briefly survey before evaluating its premises.

1. The aim of Rawls's difference principle is to maximally benefit the least well-off.
2. If a person is able to produce a benefit for the least well-off with an incentive, he can produce that same benefit without an incentive.

Sub-conclusion: Incentives are not objectively necessary to maximally benefit the least well-off.

3. Foregone incentives to producers of benefits can be redistributed to the least well-off, further benefitting them.

Sub-conclusion: Realizing the aim of the difference principle prohibits providing producers with incentives.

4. If incentives are, subjectively, necessary for producers to act to benefit the least well-off, it is only because producers make incentives necessary (in the same way that a kidnapper makes paying a ransom necessary).
5. If citizens internalize and act on an egalitarian ethos, producers do not make incentives necessary.

Therefore, realizing the aim of the difference principle requires that citizens internalize and act on an egalitarian ethos.[5] If we then make the plausible assumption that acting to gain a proffered incentive is acting on self-interest, whereas (for Cohen) acting on an egalitarian ethos is exhibiting citizen virtue, we have an argument that civility requires not acting on self-interest by demanding productive incentives.

Cohen presented his reasoning as an argument that Rawls's theory of justice is internally inconsistent because Rawls defends the use of incentives to increase productive activity, whereas the difference principle condemns their use: "It is evidently supposed by Rawls himself, that his affirmation of the difference principle is consistent with his endorsement of the inequalities that come with special incentives to people of talent. But I shall argue that, when true to itself, Rawlsian justice condemns such incentives."[6] According to Cohen, the difference principle permits inequalities in people's holdings only if they are necessary to improve the position of the least well-off[7] because the aim of the difference principle is to maximally benefit

[5]These sub-arguments appear in the works cited in note 1.
[6]Cohen, "Incentives," 378.
[7]Cohen, "Action," 8; Cohen "Incentives," 379.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

the least well-off. Incentives are offered to the "talented" to induce them to act in ways that produce benefits for the least well-off. But if the "talented" can produce benefits for the least well-off with incentives, surely, they could produce those same benefits without incentives. That they can now do it for the prospect of a bonus shows that they are now *able* to do it; if they are able to do it, then they can do it even in the absence of a bonus. Indeed, their foregone incentive payment could then be redistributed to the least well-off, further improving their situation. Hence incentives cannot be objectively necessary to improve the position of the least well-off. Therefore, the difference principle must condemn them. It follows that Rawls's theory is inconsistent in allowing them.

Moreover, if incentives are subjectively necessary to elicit productive behavior from the talented, in the sense that without them those talented will elect not to produce the benefits, then they are necessary only because the talented make them necessary. This is tantamount to extortion. Just as the kidnapper "makes it necessary" to pay ransom for the return of your child, so the talented "make it necessary" to offer them incentives for producing benefits to others. (Cohen's analogy to the morally impermissible behavior of kidnapping suggests that he sees requiring incentives for productive behavior as likewise immoral.) What is necessary is merely a matter of the producer's attitude.[8] This shows, Cohen concludes, that the difference principle's aim of maximizing the position of the least well-off cannot be realized unless the difference principle is supplemented by an egalitarian "ethos."[9] This ethos would replace producers' self-interested motivation, so that they would not demand benefits as a condition of producing benefits for others. Rawls is therefore wrong to believe that any system of markets and tax schemes could realize justice; a system of pure adjusted procedural justice among "self-interested market maximizers" cannot realize justice. Effective implementation of Rawls's principles of justice therefore requires virtuous citizens to act on an egalitarian ethos that overrides self-interest when making personal choices about productive behavior.

2. Synchronic Personal Application of the Difference Principle and its Reciprocity Requirement

Cohen's argument takes it for granted that the difference principle is suited to assess the justice of distributions of holdings between determinate persons or groups of determinate persons at a time; we can use it to criticize person A's larger holdings if those holdings exceed the amount that would be necessary at that time to maximize person B's holdings. (This synchronicity shows in

[8]Cohen, "Action," 9; "Incentives," 355–61.
[9]Cohen, "Action," 10; "Incentives," 383.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

premise 2 of that argument.) Let us call such an application of the difference principle a synchronic personal application, and provisionally assume, with Cohen, that such applications are unobjectionable.

How can one person's productivity produce economic benefits for other people? Depending on the relationship between persons A and B, B might benefit economically from A's increasing the number of hours A works, or A's increasing the intensity with which he works (his productivity per unit of time), or from A's investing his resources in productive enterprises rather than consuming them. If in a nightclub, the waiters and restroom attendants receive larger tips when Zelda sings than when others perform, they would benefit economically were Zelda to increase the number of shifts she works. If the wages of workers at the poultry processing factory depend on how many chicken nuggets the factory produces in a month, his fellow workers would benefit were Cosmo to intensify his labor throughout their standard shift, thereby producing more nuggets per unit of time. And if Joe would have to pay less tax were Nelson to invest his resources in productive enterprises rather than consume them, then Joe benefits when Nelson invests.

In all these cases, it would economically benefit the waiters and restroom attendants, factory co-workers, and Joe, to have Zelda, Cosmo, and Nelson work longer, work harder, or invest rather than consume, no matter what the effects on the economic well-being of Zelda, Cosmo, and Nelson. If Zelda works twice as long for an additional wage that just equals the cost of her childcare during those additional shifts, she does not improve her economic well-being; but the increased tips of the waiters and restroom attendants do make them better off. If Cosmo produces twice as fast, but his 1/nth share of the higher wage pool just covers the cost of the weekly chiropractic adjustment he needs because he's working so intensely, then he receives no net benefit from the higher wage even though his co-workers do. And if the scheme of taxation is such that Nelson only just recovers the cost of his investment, while its returns in excess of this go to reducing Joe's tax bill, or are actually redistributed to Joe, then Joe benefits even when Nelson doesn't. When in such situations one behaves productively, one's activity may economically benefit others without economically benefitting oneself.

Does the difference principle require Zelda, Cosmo, and Nelson to increase their productive efforts so as to maximally benefit the least well-off, even when doing so does not benefit themselves? Consider what Rawls writes: "the difference principle expresses a conception of *reciprocity*. It is a principle of *mutual* benefit";[10] "the difference principle appears to be acceptable *both to the more advantaged* and to the less advantaged individual";[11] "[on the

[10]John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 102, emphases added.

[11]Rawls, *Theory of Justice*, 104, emphasis added.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

upward rising part of the contribution curve] the *criterion of mutual benefit* is always fulfilled … only *reciprocal* advantages are allowed."[12]

If Cohen were right that the aim of the difference principle is to maximally benefit the least well-off (premise 1), the difference principle would mandate productive contributions that unilaterally advantage them. But Rawls says that the difference principle is a principle of reciprocal advantage or mutual benefit, and that "the difficulty is to show that [the *more* favored] has no grounds for complaint."[13] Hence any program that merely repaid to the more advantaged the costs of their producing benefits for others, not providing them a benefit but leaving their level of economic well-being unchanged, would not satisfy the difference principle. In this sense, the synchronic personal difference principle doesn't just permit incentives; it requires them if the elective productive efforts of the advantaged are to be reciprocally, that is mutually, beneficial. This suggests that the real aim of the difference principle is something other than to maximally benefit the least advantaged, and that Cohen's conception of "more advantaged" producers differs from Rawls's.

3. Retargeting Cohen's Argument

Upon seeing that Rawls's difference principle is in fact a principle of reciprocal benefit, Cohen could try to adjust his argument to take aim not at an alleged incompatibility between the use of incentives and the difference principle but at the difference principle itself. He could argue that Rawls's justification for adopting the principle (as a principle of reciprocal benefit) is that incentives are necessary to motivate producers to act to benefit the least well-off. If the sole (or primary) reason for adopting a principle of reciprocal benefit is that it is necessary for benefitting the least well-off, then Cohen's argument that identical benefits could be produced without incentives will, if sound, discredit the difference principle itself.

To evaluate this suggestion, we need to look at Rawls's argument for the difference principle as against a principle of strict equality. That requires looking at the argument from the original position because, Rawls insists, all arguments for principles of justice must be made from within the original position. The issue then is whether Rawls argues that the parties accept the difference principle because they think that incentives are necessary to improve the position of the less well-off, or whether he argues for the difference principle on other grounds.

Rawls's official argument for the difference principle begins by specifying the motives and interests of the parties.[14] He stipulates that the parties are

[12]Rawls, *Theory of Justice*, 104, emphasis added.
[13]Rawls, *Theory of Justice*, 103, emphasis added.
[14]Rawls, *Theory of Justice*, section 25 "The Rationality of the Parties."

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

mutually disinterested, and hence not motivated by benevolence. (He says this stipulation emphasizes the subjective circumstances of justice.) He stipulates that they are not moved by envy. And he stipulates that the parties "prefer more primary social goods rather than less."[15]

> the persons in the original position try to acknowledge principles which advance their system of ends as far as possible. They do this by attempting to win for themselves the highest index of primary social goods, since this enables them to promote their conception of the good most effectively whatever it turns out to be. The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor. Nor do they try to gain relative to each other; they are not envious or vain. Put in terms of a game, we might say: they strive for as high an absolute score as possible[16]

With these assumptions in place, Rawls argues for the difference principle as follows: If a principle of justice would allow inequalities that raise everyone's absolute score, each of the parties prefers it. The difference principle does allow such inequalities; a principle of strict equality does not: "If there are inequalities in the basic structure that work to make everyone better off in comparison with the benchmark of initial equality, why not permit them?"[17] This argument does not rely on the premise that incentives are necessary to produce benefits for the poor, nor does it offer that they should be allowed for that reason. It simply says that if there are departures from equality that make everyone better off in the long run, then the parties will select a principle that allows them. Indeed, "it would be shortsighted of [them] not to do so."[18]

We can see that this argument does not depend essentially on incentives by seeing how the parties might choose the difference principle even though the idea of productive incentives never occurs to any of them. Suppose they can imagine that directing a greater than equal share of the primary good of income to pregnant women would significantly decrease infant mortality and ongoing disability among the future citizens upon whose productive efforts we must all depend for our own economic security as we age. They are not imagining that women are being provided incentives to become pregnant; rather, they are imagining that extra income during pregnancy in the form of more nutritious foods and better medical care will improve fetal health in a way beneficial to each of the parties in the long run. If Rawls's reasoning shows, based on cases like the one imagined, that the parties will select the difference principle—even though the idea of incentives for production never occurs to them—then the need to provide incentives cannot be essential to the justification of the difference principle.

[15]Rawls, *Theory of Justice*, 142.
[16]Rawls, *Theory of Justice*, 144.
[17]Rawls, *Theory of Justice*, 151.
[18]Rawls, *Theory of Justice*, 151.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

However, that is not to say that incentive considerations must be irrelevant. We can easily imagine how incentives might come up in the parties' discussion in the original position. When presented with the difference principle, some might ask, "How is it possible that departures from equality could benefit everyone? After all, there's a fixed stock of goods, so how could dividing it unequally give everyone more than an equal distribution would?" And in response, talk of incentives could have a place. Other parties might answer, "You're assuming that the stock of goods is fixed. But suppose it isn't. Maybe some people would produce more, increasing the total stock of goods, if they were allowed to keep a greater share of the excess their extra efforts produced. Maybe they would undertake to develop and train socially beneficial abilities rather than follow their personal preferences if we advertised some special reward for doing so. If that were true, it would be possible for everyone to benefit from allowing inequality." But the role of this incentives-consideration is part of an answer to a question that's likely to arise for the parties about the conceptual coherence of the proposed alternative to strict equality.

Because the parties are stipulated not to be moved by a desire to benefit others or by the promptings of envy, we would not expect their deliberations to include consideration of such questions as "if people could be more productive, shouldn't they do so as a public service without asking what's in it for themselves?" or "shouldn't people choose which talents to develop and which life plan to pursue using altruistic criteria?" or "even if we'd all be better off, won't it be upsetting if others have more than oneself?" But they might be concerns for us, and Rawls would address them by reminding us of how the parties are characterized and of the philosophical justifications for so characterizing them. When we look at section 26 of *A Theory of Justice*, titled "the reasoning leading to the two principles of justice," we find what we were expecting. Rawls argues that, if inequalities benefit all, then it is rational for the parties to allow them. He mentions incentives in answering the suppressed question of how such a scheme could be possible; and answers two imagined objections—one from the requirements of benevolence and one from envy—reminding the reader of his stipulations and their justifications.[19]

Rawls concludes that allowing mutually beneficial inequalities doesn't differentiate among various such arrangements. "In order to make the principle regulating inequalities determinate, one looks at the system from the standpoint of the least advantaged representative man. Inequalities are permissible when they maximize, or at least all contribute to, the long-term expectations of the least fortunate group in society."[20] Thus, to the question whether Rawls's argument for the difference principle is that a principle of reciprocal benefit is necessary, given greedy human nature, in order to benefit

[19]Rawls, *Theory of Justice*, 151.
[20]Rawls, *Theory of Justice*, 151.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

the least well-off, the answer must be "no"; and so, reformulated as a criticism of Rawls's justification for the difference principle itself, Cohen's incentives argument fails. Let us return to our investigation of Cohen's argument that realizing the difference principle's aim requires rejecting incentives to producers.

4. Incentives, Super-Incentives, and Extortion

Cohen calls those able to command an incentive for their producing benefits for others "the talented rich," "self-seeking highflyers," and "high-flying market maximizers," acting on self-interest. If Zelda can command an incentive to take on more performances, that may be because of her *talent*, a talent not shared by the waiters and restroom attendants. Similarly with Cosmo's talent of producing so much more efficiently than his co-workers. If Nelson can command an incentive to invest rather than to consume, it may be because of his situation, which enables him to satisfy his needs while still having money left over to invest.

People like Nelson can vary their productive activity in light of their expected return, and they can hold out for terms. Perhaps this is what Cohen is trying to call attention to in likening such folk to "extortionists" (premise 4). Their talent or situation enables them to vary their returns with their effort, and they can hold out for favorable terms, because they do not need income urgently. So let us say that one is "well-situated" if she doesn't need to maximize her earnings and she can vary her earnings with her economic activities, adjusting her efforts to her needs as they arise. She has marketable skills which are in demand, and so can increase her earnings by prolonging her work; or she has talents that make it possible for her to switch fields or retool so as to exploit earning opportunities. Or she has resources to invest, and so can take advantage of investment opportunities. Or all the aforementioned. And she is under no compulsion of dire need to maximize her earnings; she can afford to hold out for favorable terms. Let us say a person is "less well-situated" if he needs to maximize his earnings or cannot vary his earnings with his effort. This is a disjunctive condition. And let us say a person is "badly situated" if he both cannot earn more by his efforts and needs to maximize his earnings. Both conditions for well-situatedness fail.

Suppose we are talking about activities that would produce reciprocal economic benefits, as the difference principle requires. Is it morally objectionable (as Cohen's extortion charge implies) that the well-situated can decline to undertake activities that would economically benefit both the badly situated and themselves, if they decide "it's not worth it" to them to do the activity for the incentive offered? Say Nelson's aim is to provide his daughter with a college education. If an investment would yield a return large enough to pay her way to college, then it would be worth it to him to

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

invest. He is perfectly willing to pay the entire costs of her education if his investment can earn enough to do that; but any return less than the full cost of her education might compromise her eligibility for financial aid, subjecting her to the prospect of burdensome college debts. Unless he can be sure of a return sufficient to pay her way, he would rather take the money he would have invested and spend it on goods that improve her chances of winning a merit scholarship to college, for example, a SAT preparation course, a collection of the classics of Western literature, a math tutor, and a writing tutor. Thus, the return that makes an investment "worth it to him" depends on what is required to advance his conception of the good. Or take Zelda, the torch singer, who will only increase her work hours at the club, thereby making her unable to continue donating her amateur bookkeeping services to her church, if doing so would earn her enough money to pay a professional to keep her church's books. The question of "what's worth it" to her makes essential reference to her particular values and ends.

Well-situated people may refuse to do what would improve the economic well-being of others unless it also advances their own life plans. Does that make them, to use Cohen's terms, "self-seeking high flyers" and "high flying market maximizers"? Or "extortionists" like the kidnapper, demanding payment for their morally impermissible actions? What the well-situated are well-situated to do is to pursue their conception of the good and to advance the values given them by their comprehensive doctrines. These are not morally impermissible activities *per se.* Unless one's specific ends and means are themselves morally impermissible, there is nothing illegitimate in trying to advance one's own life plan, *ceteris paribus.*

To judge the moral significance of situatedness within Rawls's theory of justice, we need to be clear about the social world at issue. In the well-ordered society of justice as fairness there are no badly situated people because no one needs to maximize his market earnings in order to meet his basic needs. Rawls's theory is said to apply only to, and so assumes, a modern industrial society with a developed economy, in which all basic needs can be satisfied. The opening passage of *Political Liberalism* states that the first principle of justice "certainly must" be presumed to "be preceded by a lexically prior principle requiring that citizens' basic needs be met."[21] The inequalities contemplated by the difference principle are all above the "basic needs meeting" standard and whatever level beyond that comprised by an equal distribution of income and wealth. Hence no one needs to maximize his earnings to meet basic needs.[22]

[21]John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 7.

[22]Rawls sets aside the issue of the special needs of the thalidomide child or the adult who becomes quadriplegic. All cases of "special needs," understood as conditions that render normal participation in social cooperation over a complete life impossible, are to be dealt with after the general theory for the standard case is worked out.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

The well-ordered society will contain some less well-situated people who cannot vary their incomes with their efforts, or at least not much. There may be no further demand for their skills, or there may be too many competitors who possess the same skills, so that no overtime work is available. They may be unable to exercise skills they have because they are ill or temporarily disabled due to accident or imprudence. In the just society, the significance of the fact that some are less well-situated will depend on the requirements of their conceptions of the good and of realizing their values as given by their reasonable comprehensive doctrines. The less well-situated person with an inexpensive conception of the good (one that is not costly to satisfy) is no worse off for practical purposes than his well-placed counterpart. He cannot maximize his earnings, but he doesn't need more than he is able to get. Conversely, the well-situated person who finds herself with a conception of the good so costly to realize that no effort she could make would earn enough to do so, or so costly to realize that she must maximize her earnings, is no better off for practical purposes than the less well-situated person.

This has the consequence that inferior situation does not automatically command the attention of justice. Rawls famously insists that citizens are to be assigned responsibility for the content of their life plans, are to be expected to adjust those life plans to the means they can reasonably expect to enjoy under a just social system, and are to limit their claims against fellow citizens accordingly. He writes, "That we can take responsibility for our ends is part of what free citizens may expect of one another … we must assume that citizens can regulate and revise their ends and preferences in the light of their expectations of primary goods."[23] Still, if the well-situated (advantaged) are allowed to vary their productive contributions according to the standard given by their own life plan of "what's worth it" to themselves, what is to keep them from inflating their demands far beyond the minimum that would truly be needed to induce them to make the productive choice?

It does not matter what position we take on the question of what a person may permissibly demand for his productive efforts because the question reflects a faulty starting point, namely the assumption we only provisionally granted earlier that "synchronic personal applications" of the difference principle are legitimate from the standpoint of Rawls's theory of justice. Those applications treat the difference principle as appropriate for judging the justice of the distributive shares, at a time, of particular persons or groups of particular persons—as appropriately answering questions like: "Does now allowing Nelson this return on this investment maximally improve Joe's present situation?"

[23]Rawls, *Political Liberalism*, 185–86.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

5. The Diachronic Positional Difference Principle Preserves Background Justice

Cohen suspected that Rawls would not approve such applications of the difference principle, and imagined him making "the basic structure objection,"[24] namely, that "the difference principle is, by stipulation and design, a principle that applies only to institutions (to those, in particular, which compose the basic structure of society), and, therefore, not one that applies to the choices, such as those of self-seeking high flyers, that people make within such institutions."[25] According to Cohen, Rawls says that his principles of justice apply to basic institutions because basic institutions have a "profound impact on people's lives."[26] Cohen replies to his imagined "basic structure objection" to his argument for the necessity of an internalized ethos that Rawls is inconsistent to exclude personal choices from the purview of principles of justice, first, because institutions are comprised of personal choices, so principles applying to institutions must also apply to the personal choices that comprise them; and second, because Rawls's criterion for including anything within the purview of justice (within institutions of the basic structure) is that it have a profound effect on people's lives, which personal choices may well have. This "fatal ambiguity in Rawls's specification of the basic structure" is, Cohen maintains, the "major fault line in the Rawlsian architectonic"[27] and "shipwrecks not only the basic structure objection but also the whole approach to justice that Rawls has taught so many to pursue."[28]

However, Rawls wouldn't object in the way Cohen imagines because Rawls holds neither of the views Cohen attributes to him. For Rawls, principles of justice do not apply to institutions, in the sense of governing or ordering their internal workings. They apply to the interactive system —"how they fit together into one unified system of social cooperation from one generation to the next" or "the framework"—of which those basic institutions are elements.[29] For instance, the market is an institution belonging to the basic structure, but the two principles do not internally order markets, taking the place of the law of supply and demand. The Supreme Court is an institution of the basic structure, but it is not to decide individual cases affecting the wealth of the litigants by applying the difference principle so as to select the decision that maximally benefits the poorer party. Principles of justice impose constraints on the institutions belonging to the basic

[24]Cohen, "Action," 10–11.
[25]Cohen, "Action," 5,
[26]Cohen, "Action," 21.
[27]Cohen, "Action," 15.
[28]Cohen, "Action," 11.
[29]Rawls, *Political Liberalism*, 11.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

structure, with, for instance, the equal basic rights and liberties principle constraining markets by legally prohibiting the sale of humans, or political offices, or votes—but they do not directly order their internal life. So even if personal choices, rather than rules and procedures, were constitutive of institutions, it wouldn't follow that the principles of justice apply to personal choices.

Nor does Rawls espouse a "profundity of effect on people's lives" criterion for inclusion in the basic structure. In "The Basic Structure as Subject," included in *Political Liberalism*, Rawls insists that the task of the basic structure is to establish and preserve "background justice," the "background against which the activities of individuals and associations take place."[30] We require "special institutions to preserve background justice, and a special conception of justice to define how these institutions are to be set up."[31] Given these and half a dozen more equivalent descriptions of "the distinctive role of the basic structure", we can safely infer that the basic structure includes whatever fundamentally contributes to the task of preserving background justice. So even though personal choices can have "profound effects" on people's lives, it does not follow from that fact that they belong to the basic structure. The relevant question is whether personal choices contribute fundamentally to the task of preserving background justice, and to this question, Rawls's answer is unambiguously negative: Individuals cannot possibly hope to secure background justice through even their most conscientious personal efforts. Rawls writes,

> The fact that everyone with reason believes that they are acting fairly and scrupulously honoring the norms governing agreements is not sufficient to preserve background justice … the conditions necessary for background justice can be undermined, even though nobody acts unfairly or is aware of how the overall result of many separate exchanges affects the opportunities of others. There are no feasible rules that it is practicable to require economic agents to follow in their day-to-day transactions that can prevent these undesirable consequences … The role of the institutions that belong to the basic structure is to secure just background conditions against which the actions of individuals take place.[32]

Rawls explains the inadequacy of even well-meaning individuals to preserve background justice in terms of the complexity of transactions, the enormous amount of information needed correctly to apply rules, the transaction costs attendant to bargaining with many widely scattered third parties, and the constraints of practicality, publicity, and burdensomeness on the directives of justice to individuals. He offers the case of property bequests: "It is obviously not sensible to impose on parents (as heads of families) the duty to adjust their

[30]Rawls, *Political Liberalism*, 266.
[31]Rawls, *Political Liberalism*, 267.
[32]Rawls, *Political Liberalism*, 266–67.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

own bequests to what they estimate the effects of the totality of actual bequests will be on the next generation, much less beyond."[33] This same reasoning speaks against Cohen's suggestion that, to be virtuous, individuals must aim to maximize the position of the least well-off. Individuals are in no position to judge the long-run effects of their actions on the least well-off, no matter how good their intentions. Thus, an egalitarian ethos cannot be expected to contribute to the justice of background social conditions, and so will fail to meet Rawls's criterion for inclusion as an element of the basic structure.

However, Rawls does have an objection to Cohen's ethos for personal choice, which reveals his diachronic and positional, as opposed to synchronic and personal, application of the difference principle. For Rawls, the task of justice is to decide the terms on which being born into one social position rather than another is to influence citizens' life prospects. It is a fact of life that being born into some social positions does advantage one; "it seems likely," Rawls writes, "that this will be true even when the social injustices which now exist are removed."[34] Even the just society will have some who enjoy greater wealth and positions of greater responsibility and prestige than do others. Children born to such families will likely have enhanced prospects for enjoying similar advantages, as parents invest their resources in their children. Hence, Rawls insists, the aim of the difference principle is "not to eliminate contingencies from social life, for some contingencies are inevitable"[35] but rather to answer the question, "What can possibly justify this kind of initial inequality in life prospects?"[36] His distinctive thought that is that it could be justified if it could be made to be advantageous to all, and thus acceptable to all as free and equal citizens. The basic structure, when ordered by Rawls's two principles, does just that. "Those who have gained more than others … do so on terms that improve the situation of those who have gained less."[37]

Rawls clarified his idea of the least-advantaged over time. He replaced earlier characterizations as those belonging to the class of unskilled laborers, or those having less than half the median income and wealth, with those whose lifetime reasonable expectations seen from their social position are lowest. The least advantaged occupy the social position with the lowest lifetime expectations for the primary goods of income and wealth:

> The individuals who belong to the least advantaged group are not identifiable apart from, or independently of, their income and wealth … the term "the least advantaged" is not a rigid designator … Even

[33]Rawls, *Political Liberalism*, 268.
[34]Rawls, *Theory of Justice*, 78.
[35]Rawls, *Political Liberalism*, 283.
[36]Rawls, *Theory of Justice*, 87.
[37]Rawls, *Political Liberalism*, 282.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

> supposing … that … the least advantaged, identified by income and wealth, include many individuals born into the least-favored social class of origin, and many of the least (naturally) endowed and many who experience more bad luck and misfortune … nevertheless those attributes do not define the least advantaged.[38]

The difference principle pronounces that the greater expectations of the more advantaged are justified only if decreasing them would lower the long-term expectations of those occupying the least well-off position. If the greater class expectations of those who inherit advantage stimulate their training of talents that might otherwise have gone undeveloped, and the deployment of those talents in socially beneficial enterprises that might otherwise not have been pursued, thus improving over time the position of those who occupy society's bottom economic position, then society has made appropriate use of the structure of inherited advantage. It has made it advantageous to all, and so (from the point of view of the original position, which is intended to model the position of free and equal citizens) acceptable to all.

The difference principle assesses the justice of the structure of advantage that attaches to inherited social positions diachronically. "We cannot tell." writes Rawls,

> by looking only at the conduct of individuals and associations in the immediate (or local) circumstances whether, from a social point of view, agreements reached are just or fair. For this assessment depends importantly on … underlying social conditions … extending backward in time and well beyond any limited view.[39]

If the superior prospects attached to favored class positions work over time to raise the position of the least advantaged, they are to be admitted. To now renege on the legitimate expectations of those who have trained and deployed the needed abilities would be not only unfair, but imprudent, because it would discourage those presently making decisions about how to train their children's talent from directing that development toward socially desirable enterprises, thus reducing the prospects of future generations of the less well-off.

To insist, as Cohen does in premise 2, that those now trained could benefit others without collecting any benefit for doing so does not entail that they could have trained the skills they now employ without the prospect of personal benefit. Indeed, they might never have done so. Their parents might have forbidden it, saying, "learn yourself a trade that pays"; just as most parents now discourage their children from pursuing the life of the starving artist (or the philosopher). If no trade particularly paid, parents might have

[38]Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001) 59, n. 26.

[39]Rawls, *Political Liberalism,* 266–67.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

encouraged their children to do what they enjoy, unhindered by consideration of broader social benefits. And if society now reneges on the deal, as Cohen proposes it should, saying "now that you've got those doctoring skills, you must practice them at minimum wage," those least well-off at present may benefit from cheaper healthcare, but the expectations of the least well-off of the future may well decline because of current citizens' diminished willingness to train socially beneficial doctoring skills.

It is thus an important feature of Rawls's theory that it views society as an ongoing cooperative enterprise, from one generation to the next, into the indefinite future. The difference principle is not designed to assess the justice of the holdings of particular individuals synchronically. In fact, "a distribution cannot be judged in isolation from the system of which it is the outcome or from what individuals have done in good faith in the light of established expectations."[40] Moreover, there is no way of knowing what we might have been or done had our expectations been different, because "abilities and talents cannot come to fruition apart from social conditions, and as realized they always take but one of many possible forms … Developed natural capacities are always a selection, a small selection at that, from the possibilities that might have been attained."[41] Rawls concludes that "our realized abilities and talents reflect, to a large degree, our personal history, opportunities, and social position. There is no way of knowing what we might have been had these things been different."[42]

This commonsense observation makes a great deal of difference to the plausibility of Cohen's argument that virtue must replace self-interest in personal choices in order to realize Rawls's just society. For that argument even to reach Rawls's difference principle—the real, diachronic positional difference principle—Cohen needs the premise that the well-situated whose activities produce benefits for others could produce those same benefits without having been allowed the advantages they in fact enjoyed. But if Rawls is right, we have no reason to accept that premise.

6. Making Inequality Acceptable to Free and Equal Citizens

I have argued that the difference principle is Rawls's answer to his fundamental question "By what principle can free and equal moral persons *accept the fact* that social and economic inequalities *are* deeply influenced by social fortune, and natural and historical happenstance?"[43] Cohen's criticism of Rawls's theory evidences a desire to *change* these facts rather than to find a

[40]Rawls, *Theory of Justice*, 88.
[41]Rawls, *Political Liberalism*, 270.
[42]Rawls, *Political Liberalism*, 270.
[43]Rawls, *Political Liberalism*, 281, emphases added.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

way to accept them, reflecting Cohen's position, following Ronald Dworkin, that people's possession of goods ought not depend on anything they cannot choose. Cohen writes: "(T)here is injustice in distribution when inequality of goods reflects … forms of lucky and unlucky circumstance."[44] In a subsequent paper,[45] he maintained that justice requires equalizing "access to advantage," advantage being "a heterogeneous collection of desirable states of the person" produced by goods, which he terms "midfare."[46] Social justice demands that one's share of desirable existential states—of being nourished, healthy, comfortable, educated, influential, etc.—must not depend on factors one cannot choose such as native endowment, good or bad brute luck, historical circumstance, or inherited social position. "Differential advantage is unjust save [except] where it reflects differences in genuine choice"; "if there is no such thing [as genuine choice], because, for example, 'hard determinism' is true, then all differential advantage is unjust."[47] If a person's unchosen disadvantage cannot be eliminated, society must compensate him for it. Cohen's egalitarian theory of justice is what Rawls would call a "comprehensive egalitarianism" for its reliance on controversial moral and philosophical assumptions that reasonable citizens may reasonably reject. It is arguably a conception of cosmic justice, to be prosecuted by political societies, rather than a conception of social justice.

Rawlsian social justice has a narrower scope. We are not to right the wrongs of the universe—"the difference principle is not the principle of redress. It does not require society to try to even out handicaps, as if all were expected to compete on a fair basis in the same race."[48] Instead, there will be permissible inequalities in individuals' life prospects, and the role of a theory of social justice is to regulate those inequalities. In Rawls's words,

> The basic structure most likely permits significant social and economic inequalities in the life prospects of citizens depending on their social origins, their realized natural endowments, and the chance opportunities and accidents that have shaped their personal history. Such inequalities, we may assume, are inevitable, or else highly advantageous in maintaining effective social cooperation. Presumably there are various reasons for this, among which the need for incentives is but one … What the theory of justice must regulate is the inequalities in life prospects between citizens that arise from social starting positions, natural advantages, and historical contingencies.[49]

[44]Cohen, "Action," 12.

[45]G. A. Cohen, "Equality of What? On Welfare, Goods and Capabilities," *Recherches Economiques de Louvain/Louvain Economics Review* 56, no. 3–4 (1990): 357–82.

[46]Cohen, "Equality of What?," 380.

[47]Cohen, "Equality of What?," 381.

[48]Rawls, *Theory of Justice*, 101.

[49]Rawls, *Political Liberalism*, 270–71.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

In Rawls's view, social justice concerns which choice we as a society make of the *terms* upon which those born into or otherwise occupying social positions and with talents that contribute to their being "well-situated" to realize their ends and values are to enjoy those advantages. Are they to enjoy them at the expense of others? Or regardless of the fate of others? Or are they to enjoy them only on terms that also benefit others? The difference principle provides an answer to this question.

Conclusion

Cohen's argument faltered at several points. First, it mischaracterized the aim of the difference principle. The aim of the difference principle is to make the inevitable array of differentially advantaged social positions acceptable to all as free and equal citizens, not to maximally benefit the least talented and unluckiest individuals in society. For Rawls, the least advantaged social position is picked out by the lowest lifetime expectations for the social primary goods of income and wealth. Second, the fact that a person at a time is able to do an action while foregoing an incentive does not suffice to show that raising the level of the least advantaged position in society does not require a long-term system of incentives to develop socially beneficial capacities, along with the honoring of legitimate expectations. Even accepting Cohen's discredited assumption that synchronic personal applications of the difference principle are legitimate (which we should not), incentives might well be objectively necessary for achieving the reciprocal benefit the difference principle requires.

Additionally, deciding whether to undertake economic activity by consulting one's own values and conception of the good is not morally analogous to a kidnapper's extorting a ransom because, unlike kidnapping, pursuing our permissible values and conceptions of the good is a legitimate activity for citizens. Moreover, Rawls insists that an ethos of justice to inform personal choices would not suffice nor even help to secure background justice—which is the purpose of regulating society's basic structure by principles of justice—so there would be no point in requiring citizens to forego pursuit of their legitimate personal interests to advance justice. For these reasons, the prospect is dim for developing any successful argument to show that political virtue in Rawls's just society requires overriding self-interest in personal choices; but in any case, Cohen's particular argument manifestly fails to show that.

I conclude by highlighting two attractive features of Rawls's system that we can now see more clearly. For Rawls, justice is the first virtue of social institutions, but not necessarily of individuals. Individuals need not make pursuit of social justice their alpha aim in all their actions. In this way Rawls's just society is much less demanding—its "strains of commitment"[50] are less

[50]Rawls, *Theory of Justice*, 176.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press

—than Cohen's comprehensive egalitarianism. Background justice in Rawls's pure (adjusted) procedural system *liberates* citizens to form, revise, and to pursue their permissible conceptions of the good in line with their values, talents, tastes, and circumstances. Rawls accepts that the freedom secured by a liberal society will result in a continuing pluralism of reasonable comprehensive doctrines and conceptions of the good, with no single dominant value shared by all citizens, including Cohen's equalization of midfare.

Second, and relatedly, by not seeking to replace such a fundamental human motivation as self-interest with a disinterested moral motive for personal decision-making, as Cohen does, Rawls's system belongs to the historical camp of political philosophies that seek to channel rather than to change human nature. I think that is a plus. A philosophy like Rousseau's, which requires citizens to decide questions affecting other citizens by consulting only their general will rather than their private or corporate wills, demands a radically altered (improved) human nature, propped up by an arbitrary civil religion and a censorship to keep citizens virtuous after an elaborate educational system forges the virtuous character in them. This is a highly ambitious, probably unrealistic, and evidently illiberal social program.

In contrast, the American founders arguably aimed to devise a system that could channel existing human motivations, checking and balancing competing self-interests to sustain an environment conducive to the exercise of individual liberties. Madison's contention, that if men were angels no government would be needed, strongly suggests that he saw government as managing citizens' non-angelic (perhaps because self-interested) natures. Rawls's system takes citizens of modern industrial democratic societies as we are—rational (to the degree necessary to see a point to social cooperation), reasonable (to the degree necessary to be able to engage in fair social cooperation), with determinate ends of our own and an interest in preserving conditions for revising and pursuing those ends. His system doesn't depend on altruism, self-sacrifice, internalization of some particular, shared comprehensive religious or moral doctrine, or any other radical transformation of our motivations. Taking us as we are, and the basic structure as it might be, the well-ordered society of justice as fairness offers what Rawls could plausibly claim to be a realistic utopia.

Acknowledgments

I thank participants in the conference "Rawls's Theory of Deliberative Democracy and Its Relevance in the Digital Age" held at Universität Regensburg in March 2024 for helpful comments on a draft of this paper, especially Paul Weithman and Eva Odzuck. I am grateful to John Dreher, Ed McCaffrey, Ed McCann, and Seanna Schiffrin for discussion of an earlier version of this paper.

https://doi.org/10.1017/S0034670525000051 Published online by Cambridge University Press