



August 28, from 4-7 pm

Lester Pollock Room, FH, 9th Floor

Colloquium in Legal, Political, and Social Philosophy

Conducted by

Liam Murphy and Samuel Scheffler

Speaker: Michael Otsuka, Rutgers University

Paper: Reciprocity, Justifiability to Each, and the Difference Principle



Colloquium Website: <http://www.law.nyu.edu/node/22315>

August 14, 2025

[#1] Reciprocity, Justifiability to Each, and the Difference Principle¹

Michael Otsuka

[#2]² Rawls is regarded as an egalitarian political philosopher largely on account of his commitment to the difference principle – which calls for a distribution of income and wealth that is to the greatest benefit of the least advantaged. His endorsement of the basic liberties and their priority is associated with his liberalism rather than his egalitarianism. Yet Rawls classifies the basic liberties as an idea of equality, whereas the difference principle is one of reciprocity.³ The latter idea figures most prominently in his final, considered defense of the difference principle in *Justice as Fairness: A Restatement*. [#3] Rawls maintains there that there is a “deeper idea of reciprocity implicit in the difference principle” (§37.3), which is itself “essentially a principle of reciprocity” (§18.3).⁴ Moreover, public affirmation of the difference principle by the more advantaged “conveys to the less advantaged their acceptance of an appropriate idea of reciprocity in the clearest possible way.” (§37.3)

What is the idea of reciprocity, and does it provide a compelling justification for the difference principle? I shall show in Section II that the reciprocity-based case for the difference principle which figures prominently in *Justice as Fairness* rests upon an unjustifiable sequencing of unequal divisions of income and wealth that are to the benefit of each in comparison with an equal division. In Section III I shall explain how a natural and theoretically well-motivated response to this problem gives rise to a reciprocity-based justification of a principle of “restricted utility” rather than the difference principle. Finally, in Sections IV-VII, I shall defend a contrasting account of reciprocity which provides a sound case for the difference principle that is

¹ Acknowledgements: Frances Kamm and other participants in a Rutgers graduate seminar on this topic. Written comments: Will Combs, Tomi Francis, Frances Kamm, Kacper Kowalczyk, Nir Eyal, Bastian Steuer, Rutger van Oeveren. Presentations: UC Berkeley, University of Washington, Binghamton University, Rutgers, McGill, University of Illinois, LSE.

² These numbers are references to corresponding PowerPoint slides, which can be downloaded [via this link](#).

³ He writes: “the idea of society as a fair system of social cooperation is quite naturally specified so as to include the ideas of equality (the equality of basic rights, liberties, and fair opportunities) and of reciprocity (of which the difference principle is an example)” *Justice as Fairness: A Restatement* (Harvard University Press, 2001), §27.2.

⁴ All references in this format are to the numbered sections of *Justice as Fairness*.

more personal in nature than the sort that Rawls is willing to countenance. Such a reciprocity-based justification is sensitive – in a way that both Rawls and his utilitarian opponents are insensitive – to the moral significance of the fact that particular individuals are better or worse off than they could have been. It thereby uniquely delivers a genuine justification of the difference principle to each person. I shall begin, in Section I, with some general remarks on the nature of Rawlsian reciprocity and its relation to the difference principle.⁵

I. Reciprocity and the Difference Principle

[#4] In *Political Liberalism*, Rawls maintains that “Fair terms of cooperation specify an idea of reciprocity” involving mutual advantage arising from “a suitable benchmark of comparison”.⁶ On his preferred conception of “justice as fairness”, the “appropriate benchmark” is one “of equality”. More specifically, he maintains that the difference principle contains an “implicit reference to equal division as a benchmark”.⁷ He contrasts this with mere “mutual advantage understood as everyone’s being advantaged with respect to each person’s present or expected future situation as things are” in an unjust “society in which property, in good part as a result of fortune and luck, is very unequal”. Gains arising from such an unequal division would be tainted by asymmetries in bargaining power arising from previous injustices.⁸

[#5] When he makes the case in *Justice as Fairness* for the choice of the difference principle as satisfying reciprocity, this benchmark is specified as an equal division of the primary goods of income and wealth at the origin of Figure 1.

⁵ One might think that, by now, everything of significance which could be said has been said about Rawls on the difference principle. That would be a mistake, among other reasons because Rawls’s reciprocity-focused defense of that principle in *Justice as Fairness* has received surprisingly little attention to date – in contrast with the extensive literature on what Rawls says about the difference principle in *A Theory of Justice* (Harvard University Press, 1971 [1999 rev. ed.]).

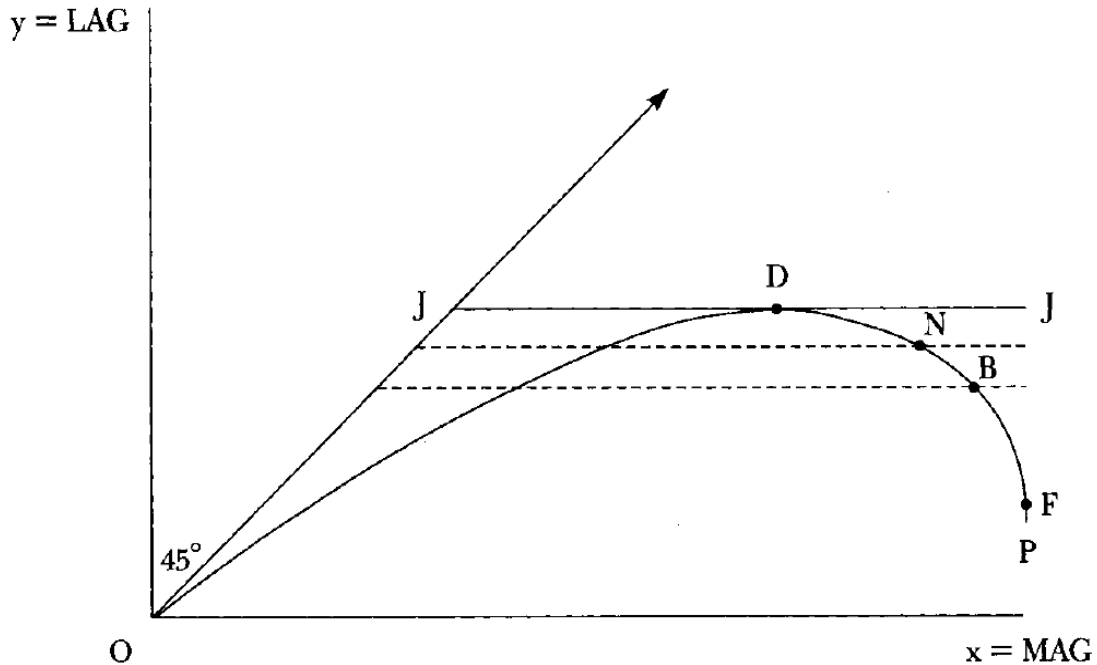
⁶ p. 16.

⁷ p. 17.

⁸ p. 17. See also *Justice as Fairness*:

the conditions for a fair agreement between free and equal persons ... must eliminate the bargaining advantages that inevitably arise over time within any society as a result of cumulative social and historical tendencies. “To persons according to their threat advantage” (or their de facto political power, or wealth, or native endowments) is not the basis of political justice. Contingent historical advantages and accidental influences from the past should not affect an agreement on principles that are to regulate the basic structure from the present into the future. (§6.2)

Figure 1



In this figure the distances along the two axes are measured in terms of an index of primary goods, with the x-axis the more advantaged group (MAG), the y-axis the less advantaged (LAG). The line JJ parallel to the x-axis is the highest equal-justice line

In Figure 1, each of the points along the curve that extends from O to P constitutes mutual advantage in comparison with this benchmark O of an equal division. These are the gains that can be achieved through socially cooperative productive labor.⁹ This OP curve has the shape it does on account of its representation of different distributions of post-tax income and wealth under different rates of taxation and transfer. The low lefthand point of the origin arises from a highly progressive scheme of taxation and transfer where, for example, pre-tax differences in earnings are taxed and transferred at a sufficiently high rate to ensure that the post-tax income of the more advantaged group (MAG) ends up no greater than that of the less advantaged group (LAG).¹⁰ Given the empirical assumption that individuals would lack incentives to develop and

⁹ “P” is so-called to stand for “production”. (§18.1)

¹⁰ Rawls maintains that the difference principle would “not involve any more continuous or regular interference with individuals’ plans and actions than do, say, familiar forms of taxation” (§14.4). However, he leaves open whether the system of taxation that realizes this principle would take the form of a progressive tax of income as opposed, say, to a progressive tax on consumption that is sensitive to level of income. See §49.4.

employ productive skills in this event, we have an explanation for the low level of income and wealth of the origin, relative to other points along the OP curve.¹¹

As we move from left to right, the points on the curve represent the distribution of post-tax income and wealth when we gradually reduce the rate of redistributive taxation of higher earners. We will eventually reach an optimal level of taxation, insofar as the post-tax income of the LAG is concerned. That is point D, which realizes Rawls's difference principle, since it maximizes the position of the LAG. A further reduction in the rate of redistributive taxation would be optimal insofar as the per capita income of all productive members of society is concerned, since it maximizes the mean income of the MAG and the LAG, when weighted by the size of the two groups.¹² This is the Bentham point B, so-called because it maximizes average utility, when utility is assumed, as Rawls does, to be linear in income and wealth.¹³ We will eventually reach the "feudal point" F which is optimal for the MAG, since the rate of taxation of higher earnings is so low that their post-tax income is maximized. Given how low this rate of taxation is, there will be little if any transfer to the LAG, and hence their level of income and wealth will be low.¹⁴

[#6] All the points on the OP curve to the left of D are Pareto-inefficient, as it is possible to move from any one of these points to another point on the curve which is strongly Pareto-superior to it since everyone is better off. For example, any adjacent point on the curve to the right of any given point to the left of D is strongly Pareto-superior to that point. By contrast, each of the points on the curve from D to F is a Pareto-efficient point: i.e., one from which it is impossible to depart by moving to another point on the curve without making at least one party worse off. A salient Pareto-efficient point in between D and F, which isn't explicitly represented on Figure 1 but which Rawls regards as providing "the strongest rival" (§34.2) to the difference principle, is that which a "principle of restricted utility" would select. This is the point that maximizes average utility (income and wealth), subject to the restriction that the least advantaged do not fall

¹¹ As Rawls writes: "By varying wages and salaries, more may be produced. This is because over time the greater returns to the more advantaged serve, among other things, to cover the costs of training and education, to mark positions of responsibility and encourage persons to fill them, and to act as incentives." (§18.2)

¹² For simplicity, Rawls assumes that society divides into just two group, the MAG and the LAG (§36.1). Hence, the "less advantaged" and the "least advantaged" co-refer, as do the "more advantaged" and the "most advantaged". I shall treat these co-referring terms as interchangeable.

¹³ See text to Figure 1 and §36.3.

¹⁴ Point N in Figure 1 is the Nash point where the product, rather than the sum, of utilities is maximized. I shall ignore this point in further discussion.

below a floor provided by a “suitable social minimum.”¹⁵ Throughout this discussion, I shall assume that B represents this point because it doesn’t fall below such a floor. I shall henceforth refer to this point as B^{RU}, to indicate that this Bentham point does not fall below the floor.

Rawls maintains that we should not settle on any of the inefficient points to the left of D, since it would be “irrational” to remain at that point when we could move to another where everyone is better off.¹⁶ Rather, we should settle on one of the efficient points. But how should we choose among the multiple efficient points from D to F? Why, for example, should we favor D over B^{RU}, as Rawls maintains we should? Rawls’s answer is that “the difference principle includes an idea of reciprocity [which] distinguishes it from the restricted utility principle” (§36.1). But how does the idea of reciprocity favor the difference principle over restricted utility? On the above characterization of reciprocity as mutual advantage against a benchmark of equality, each of these points would appear to realize this idea.

[#9] Rawls nevertheless maintains that:

D is the only point on the (highest) OP curve that meets the following reciprocity condition: those who are better off at [that] point are not better off to the detriment of those who are worse off at that point. Since the parties represent citizens as free and equal, and thus take equal division as the appropriate starting point, we say this is an (not the only) appropriate reciprocity condition. We haven’t shown there is no other such condition. But it is hard to imagine what it might be. (§36.3, my italics added)¹⁷

At every point on the OP curve, apart from the origin, there is inequality between the groups: the MAG is always better off than the LAG. But D uniquely possesses the property of there being no other point at which the worse off are at least as well off as they would be at that point.

Therefore, at any point, apart from D, those who are better off at that point in comparison with D

¹⁵ Like the difference principle, this principle of restricted utility is also constrained by Rawls’s equal basic liberties and fair equality of opportunity. **Show background slides: [#7] & [#8]** on the “second fundamental comparison”, **then return to [#6]**.

¹⁶ Rawls would deem this irrational because “it does not allow society to meet the requirements of social organization and efficiency.” (§46.1(a))

¹⁷ To capture what Rawls must have meant by this italicized claim, I have interpolated “[that]” in replacement of Rawls’s word “any”.

are better off to the detriment of the worse off, since the worse off are worse off than they would be at point D. Hence the italicized claim in the above passage is sound.¹⁸

Why, however, does Rawls find it hard to imagine any other “appropriate reciprocity condition” which a point other than D might satisfy?

[#10] In summarizing his reciprocity-based case for the difference principle, Rawls writes:

To sum up: the difference principle expresses the idea that, starting from equal division, the more advantaged are not to be better off at any point to the detriment of the less well off. But since the difference principle applies to the basic structure, a deeper idea of reciprocity implicit in it is that social institutions are not to take advantage of contingencies of native endowment, or of initial social position, or of good or bad luck over the course of life, except in ways that benefit everyone, including the least favored.... (§36.4)

The “deeper idea of reciprocity” articulated in this passage does not, however, uniquely pick out D. B^{RU} – and indeed all efficient points on the OP curve from D to (and including) F – are to the benefit of everyone, including the least favored, as measured against the “suitable benchmark” of an equal division at O. On this formulation of reciprocity, as well as the formulations from *Political Liberalism* quoted above, there is no privileging of gains to any one group (e.g., the LAG) over any other group (e.g., the MAG). We might characterize these as formulations of reciprocity which are *neutral* between the less advantaged and the more advantaged in their statement of the requirement of mutual advantage.

Neither D nor B^{RU} is to the mutual advantage of each, in comparison with the other point, though both are to the mutual advantage of each in comparison with the benchmark equal division of the origin. In order to privilege the difference principle over restricted utility, we need to appeal to

¹⁸ It is sound so long as we assume that the same individuals constitute the worse off group across different points on the OP curve. In Section V, I shall show how serious difficulties arise for the italicized claim, and therefore for Rawls’s defense of the difference principle, when we relax this assumption.

something above and beyond reciprocity in the neutral sense of mutual advantage as against a benchmark of an equal division.¹⁹

II. Rawls's ascent to peak D of the OP curve

[#11] In §36.3, Rawls provides the following argument which traces a path along the OP curve from O to D, in a manner which distinguishes D from B^{RU} and privileges the former point over the latter:

36.3. To see one way the parties might arrive at the difference principle, consider Figure 1. Imagine they have agreed to move from O to D, as everyone gains in the segment OD and D is the first (Pareto) efficient point.

At D the parties ask whether they should proceed from D to B, which is on the southeast-sloping part of the OP curve to the right of D. ... The points in the segment D to B and on to the point F [...] are also efficient points: movements along that segment can raise the index of one group only by lowering the index of the other. The segment DF is the conflict segment in contrast to the segment OD along which everyone benefits by moving northeast.

The difference principle represents an agreement to stop at D and not to enter the conflict segment.²⁰

In contrast to movement along the curve from the origin to D, movement along the curve from the origin to B^{RU} doesn't always involve mutual advantage. As Rawls notes above, D is the "first" point along the OP curve, when one moves left to right from the origin, at which no further gains are mutually advantageous. All further gains are at someone's expense and hence neither strong nor weak Pareto improvements.

[#12] Rawls maintains that, beyond D, "the reciprocity implicit in the difference principle no longer obtains". Further moves along the curve beyond this point are "[c]ontrary to reciprocity"

¹⁹ Following Rawls, my ensuing discussion of the case for the difference principle will focus on its relative merits in his "second fundamental comparison" with its "strongest rival" principle of restricted utility, similarly constrained by equal basic liberties and fair equality of opportunity.

²⁰ The passage quoted earlier, starting with "D is the only point on the (highest) OP curve that meets the following reciprocity condition", follows directly from this passage.

because they involve “trade-offs” in the “conflict segment” of the OP curve (§18.1, text to Figure 1). Rather than trade-offs of the interests of some against the interests of others, reciprocity requires moves that are mutually advantageous – i.e., strong Pareto improvements. Although points to the right of D are mutually advantageous in comparison with the origin (O), movement along the OP curve to the right of D is in violation of reciprocity if we regard the last point from which one has moved to another point as the new benchmark against which a Pareto improvement is measured. In describing this as the “new” benchmark, I am assuming that each move along the curve to the right of the origin gives rise to a new benchmark, against which further changes must be measured and determined to be mutually advantageous.²¹

As any further movement to the right of D and into the conflict segment would involve the trade-off of a gain to the better off against a loss to the worse off, Rawls maintains that the better off would express a commitment to reciprocity by agreeing to refrain from moving into this conflict segment:

[#13] since the difference principle expresses an agreement not to enter the conflict segment, and since the more advantaged, who hold positions of authority and responsibility, are better placed to enter it, their publicly affirming that principle conveys to the less advantaged their acceptance of an appropriate idea of reciprocity in the clearest possible way. (§37.3)

[#14] As I shall now show, this argument for privileging D over B^{RU} is unsound, since it rests on the arbitrary fact that D is the “first” efficient point one encounters when one moves from left to right along the OP curve, starting at the origin (O). Given that it uniquely involves an equal distribution of income and wealth, Rawls has grounds to privilege O by starting there. This is an “appropriate benchmark” of equality. But he lacks a rationale for next considering, by moving to, points adjacent to O, involving rightward movement along the OP curve. It is as if Rawls conceives O as the “starting point” of a journey involving the traversing of adjacent points on the OP curve over time, of which Figure 1 functions as the roadmap. He should instead regard O as a benchmark with which any other point on the OP curve might be compared. Once we set aside the misleading metaphor of a trip along a road that begins at point O, there is nothing to prevent

²¹ Note that, unlike the origin, these new benchmarks aren’t ones of an equal division.

an initial comparison of O with B^{RU} . Recall that the latter is the point which maximizes average utility, subject to the constraint of a suitable social minimum.²² Like D, B^{RU} is also a strong Pareto improvement over O.²³ If, having established the case for B^{RU} over O, we then treat B^{RU} as a benchmark against which other points on the OP curve, including D, are to be assessed, those points will all suffer the following deficiency, relative to B^{RU} : they are better for some, only at the expense of being worse for others.

When we recall why the OP curve in Figure 1 has the particular shape that it has, it will become clear why Rawls has no good reason to prioritize continuous movement along the OP curve from left to right starting at the origin. Recall that the different points represent higher and lower rates of taxation and the redistributive transfers which they make possible. The rationale for the shape of this curve is along lines of the rationale for the shape of a Laffer curve [#15], save for the fact that a Laffer curve is typically represented by 0% taxation rather than 100% taxation at the origin.²⁴ Note that controversy regarding the Laffer curve isn't over the claim that the endpoints involving 0% taxation and 100% taxation are low, where points along the Laffer curve represent the amount of revenue generated from different rates of taxation. What is contested is where the peak of this curve lies: i.e., how high or low the rate of taxation between these two low endpoints is, which maximizes tax revenue.

There is no reason to prioritize points to the left of the curve over points to the right because there is no reason to prioritize higher levels of taxation over lower levels, as benchmarks against which the possibility of a Pareto-improvement is to be assessed. Hence, no normative significance can be accorded to the fact that D is the "first" point along the OP curve when one travels to the right of the origin on a continuous path along the curve. This is because no rationale is offered (or can be provided) for why one must travel to that point via a continuous path along the OP curve from the origin. One must therefore make the case for D without

²² Recall that B maximizes average income and wealth, which are assumed to be linear in utility.

²³ It is also a Pareto-improvement over all points between O and the point at which the dotted horizontal line that intersects B also intersects the rising lefthand side of the OP curve.

²⁴ See John Quiggin <https://crookedtimber.org/2014/09/08/rawls-bentham-and-the-laffer-curve/> and <https://www.facebook.com/johnquiggin/posts/10152724101337386>.

privileging it as the “first” point one reaches in the “conflict segment” in a left to right journey along the curve from the origin.²⁵

[#16] Given the shape of Rawls’s OP curve, D is not just the “first”, but also the *only*, efficient point on the OP curve which is better for both the LAG and the MAG than any of the inefficient points (i.e., the points from O to D). It therefore satisfies reciprocity as involving a strong Pareto improvement, when measured against a wider range of benchmarks involving all the *inefficient* points on the OP curve.

One might try to argue that this feature is normatively significant for the following reason: although there is a presumption in favor of equality, it would be “irrational” (recall §46.1(a) cited in fn. 16) to stick to equality when we could depart from equality in a manner that is better for all. For any such strong Pareto-improvement over this equal benchmark, it would also be irrational to remain at this point, if we could move to a further point which is strongly Pareto superior to it. Moreover, D is the only point which has the following virtue in comparison with every one of these points at which it would be irrational to remain because they are Pareto inefficient: it is the only point which is strongly Pareto superior to every Pareto-inefficient point.

It’s hard, however, to see how this feature privileges D over other efficient points. Why should the fact that D is Pareto-superior to all points at which it is irrational to remain redound to D’s credit? Given this defect of irrationality, none of these points are genuine rivals to D. The fact that D possesses a unique advantage over the full range of defective points is not to D’s credit in any obvious respect. By contrast to any of the inefficient points to the left of D, at least some of the efficient points to the right of D are genuine rivals to D. We should ask what direct advantage D has over such points as B^{RU} , rather than what advantages D has over B^{RU} , insofar as their relative advantages over defective, non-rival points are concerned.²⁶

²⁵ In a short unpublished note, I reject Rawls’s claim that his assumption of “close-knitness” renders the “lexical” version of the difference principle superfluous. I show that his claim rests on the unjustified assumption that only continuous moves along the OP curve from left to right are admissible.

²⁶ In *Theory*, Rawls attempts as follows to privilege points to the left of D over those to D’s right:

...a society ... should operate only on the upward rising part of the contribution curve (including of course the maximum). On this segment of the curve the criterion of mutual benefit is always fulfilled. Moreover, there is a natural sense in which the harmony of social interests is achieved;

[#17] III. An argument from reciprocity for restricted utility

[#18] The following question remains unanswered: how does reciprocity privilege the maximal gains to the LAG that the difference principle calls for? Nozick once famously characterized Rawls's claim that the difference principle represents fair terms of cooperation between the better and worse endowed as follows: "Look, better endowed: you gain by cooperating with us. If you want our cooperation you'll have to accept reasonable terms. We suggest these terms: We'll cooperate with you only if we get *as much as possible*." He then asks us to imagine the better endowed making the "almost symmetrical opposite proposal": "Look, worse endowed: you gain by cooperating with *us*. If you want our cooperation you'll have to accept reasonable terms. We propose these terms: We'll cooperate with you so long as *we* get as much as possible."²⁷ This would involve an embrace of F (the feudal point), which is the mirror image of D. "If these terms seem outrageous, as they are," asks Nozick, "why don't the terms proposed by those worse endowed seem the same?"²⁸ Why, in light of Nozick's challenge, should we embrace Rawls's difference principle, with its apparently extreme privileging of the LAG over the MAG? How does this show reciprocity towards the more advantaged?

[#19] In *A Theory of Justice*, Rawls addresses such a critique of the difference principle, in noting that "[a]t first sight ... it may appear unfairly biased towards the least favored". In response, Rawls writes:

It seems clear that society should not do the best it can for those initially more advantaged [i.e., should not select the feudal point F]; so if we reject the difference principle, we must prefer maximizing some weighted mean of the two expectations. But if we give any weight to the more fortunate, we are valuing for their own sake the gains to those already more favored by natural and social contingencies. No one had an

representative men do not gain at one another's expense since only reciprocal advantages are allowed. (p. 89 [rev. ed.])

But (as noted above) mutual benefit is also fulfilled at B^{RU} on the downward sloping segment of the OP curve, insofar as everyone is better off there in comparison with a benchmark of equality, and also in comparison with many points on the upward sloping portion of the curve.

²⁷ *Anarchy, State, and Utopia*, p. 195.

²⁸ *Ibid.*

antecedent claim to be benefited in this way, and so to maximize a weighted mean is, so to speak, to favor the more fortunate twice over.²⁹

[#20] There is, however, a strong case that choosers in the OP from behind the veil would opt for a principle of restricted utility over the difference principle. If this case is sound, then a principle of restricted utility, which constitutes a weighted mean of the expectations of the LAG and the MAG, would be justified – not as a “valuing for their own sake the gains to those already more favored”, but rather – as in each party’s rational self-interest, when chosen in the original position. As I shall now explain, this would provide grounds of reciprocity for restricted utility.

[#21] Rawls maintains that “the principles most appropriate to specify the fair terms of social cooperation between citizens regarded as free and equal” are those “that would be selected by citizens themselves when fairly represented as free and equal” (§22.4). The original position provides such representation, since free and rational choosers are equally situated behind a veil of ignorance that deprives all of knowledge of their natural assets and socially developed talents, rendering it impossible for the more fortunate to exploit their advantages.

[#22] We pose the following question: of the two ex post efficient points D and B^{RU} which provide the main rivals to one another, would one of them be chosen over the other as in the ex ante self-interest of each of the parties in the original position? If a principle of restricted utility rather than the difference principle would be chosen by all parties in the original position, then distribution D called for by the difference principle is inefficient in the following respect: it is strongly ex ante Pareto inferior, from behind the veil, to restricted utility. It would therefore be irrational to select and remain at the difference principle, as parties would prefer restricted utility to it from the perspective of the original position. The principle of restricted utility would thereby realize the idea of reciprocity by capturing mutual gain from a benchmark of equality.³⁰ Rather

²⁹ *Theory*, p. 88 [rev. ed.].

³⁰ Rawls also defends the difference principle as providing “a natural focal point between the claims of efficiency and equality” by virtue of the fact that it is the efficient point on the OP curve which lies closest to the 45-degree line of an equal division (§36.2). But why doesn’t choice from behind the veil in the original position provide the right balancing of equality (via symmetry of the parties) and efficiency (via rational self-interest of the choice)? As noted above, Rawls maintains that “the principles most appropriate to specify the fair terms of social cooperation between citizens regarded as free and equal” are those that would be chosen in the original position (§22.4). As I have also noted, there is a strong case that parties would choose B^{RU} over D in the original position.

than an equal division of income and wealth in society, that benchmark is the original position itself within which rational self-interested choice occurs, where the parties are represented as equal on account of their placement behind the veil.³¹

[#23] It would in fact be rational for parties in the original position to choose a principle of restricted utility for the following reason. Rawls is not justified in invoking any of the three conditions that call for application of the maximin rule in his “second fundamental comparison” of the difference principle with a principle of restricted utility.³² Rather, on the reasonable assumption, which Rawls lacks good ground to reject, that a chooser from behind the veil has an equal probability of being each member of society, it would be rational for parties in the original position to maximize their arithmetic mean expectation of primary goods, rather than maximizing the minimum they might end up with. The maximization of this expectation implies the choice of a principle of restricted utility over the difference principle for the society that one inhabits.³³

[#24] IV. Alternatives to the difference principle would impose an unreasonable demand on the LAG

[#25] In the previous section I outlined an account of Rawlsian reciprocity that provides a justification for the principle of restricted utility, since that (rather than the difference principle) is what all rationally self-interested choosers would prefer when they are equally situated. Why does Rawls nevertheless maintain that reciprocity privileges the LAG and calls for the difference principle? As I noted in the previous section, one reason why is that he thinks the conditions that

³¹ See *Theory*, where Rawls downplays the significance of an equal division of income and wealth:

One obvious sense in which this is so [i.e., “if the [difference] principle is satisfied, everyone is benefited”] is that each man’s position is improved with respect to the initial arrangement of equality. But it is clear that nothing depends upon being able to identify this initial arrangement; indeed, how well off men are in this situation plays no essential role in applying the difference principle. We simply maximize the expectations of the least favored position subject to the required constraints. As long as doing this is an improvement for everyone, as so far I have assumed it is, the estimated gains from the situation of hypothetical equality are irrelevant, if not largely impossible to ascertain anyway. (p. 69 [rev. ed.])

³² **Recall background slides: [#7] & [#8]** on the “second fundamental comparison”.

³³ I provide a defense of the claims in this paragraph in Appendix A of this paper. [I have not included the appendices in this pre-read for the NYU Colloquium. However, you may download them [via this link](#).]

call for the maximin rule obtain in the original position. As I shall discuss in this section, I think another reason why is that Rawls conflates an understanding of reciprocity as mutual advantage from a benchmark of equality with a different understanding of reciprocity as mutual (or universal) acceptability consisting of what nobody can reasonably reject in a pairwise comparison of claims.³⁴ Even if a principle of restricted utility is implied by the former understanding in the manner that I have indicated in the previous section, the difference principle can be shown to follow from the latter.

[#26] License to attribute the latter Scanlonian/Nagelian reading of reciprocity to Rawls can be found in his remarks that his notion of the reasonable is “closely connected with T. M. Scanlon’s principle of moral motivation”, which he characterizes as “one of the three basic principles of his contractualism, as stated in ‘Contractualism and Utilitarianism,’ ...”. Rawls writes that “in setting out justice as fairness we rely on the kind of motivation Scanlon takes as basic [which is] the basic desire to be able to justify our actions to others on grounds they could not reasonably reject...”³⁵ [#27] This is of a piece with Rawls’s statement of the ‘criterion of reciprocity’ in his 1996 introduction to the paperback edition of *Political Liberalism*, which was pretty much his last word on the subject:

For these terms [of social cooperation] to be fair terms, citizens offering them must reasonably think that those citizens to whom such terms are offered might also reasonably accept them. Note that “reasonably” occurs at both ends in this formulation: in offering

³⁴ Rawls appears to conflate these distinct notions of reciprocity in passages such as the following:

Justice as fairness conjectures that the principles that will seem reasonable [...], all things considered, are the same principles that rational representatives of citizens, when subject to reasonable constraints, would adopt to regulate their basic institutions. What constraints, though, are reasonable? We say: those that arise from situating citizens’ representatives symmetrically when they are represented solely as free and equal, and not as belonging to this or that social class, or as possessing these or those native endowments, or this or that (comprehensive) conception of the good. (§23.3)

the representatives of citizens as reasonable and rational agents must be situated reasonably, that is, fairly or symmetrically, with no one having superior bargaining advantages over the rest. This last is done by the veil of ignorance. (*Political Liberalism*, pp. 52-53)

In notes entitled “Rawls on reciprocity and its relation to the concepts of reasonable agreement and mutual advantage”, I sketch an account of how and why Rawls associates reasonable agreement with mutual advantage.

³⁵ *Political Liberalism*, pp. 49-50, n. 2.

fair terms we must reasonably think that citizens offered them might also reasonably accept them.³⁶

[#28] Rawls's most compelling argument for the difference principle involves what I take to be an appeal to what nobody can reasonably reject in a pairwise comparison of claims. Recall the following reciprocity condition of Rawls which favors the LAG, in a manner that implies the difference principle:

D is the only point on the (highest) OP curve that meets the following reciprocity condition: those who are better off at [that] point are not better off to the detriment of those who are worse off at that point. (§36.3)

Rawls maintains that, in contrast to the difference principle, the principle of restricted utility would impose an extreme demand upon the worse off:

...in asking the less advantaged to accept over the whole of their life fewer economic and social advantages ... for the sake of greater advantages ... for the more advantaged, the principle of [restricted] utility asks more of the less advantaged than the difference principle asks of the more advantaged. Indeed, asking that of the less advantaged would seem to be an extreme demand. (§38.2)³⁷

[#30] I think we can generalize something along the lines of this objection to restricted utility into a compelling objection that *any* alternative to the difference principle involving another

³⁶ p. xlv. Cf. Rawls's much earlier characterization of "the concept of justice as reciprocity" in "Justice as Reciprocity," which was published in 1971: "persons must be regarded as possessing an original and equal liberty, and their common practices are unjust (or...unfair...) unless they accord with principles which persons so circumstanced and related could be reasonably expected to acknowledge and freely accept before one another" (*Collected Paper*, ed. Samuel Freeman (Harvard University Press, 1999), p. 192).

³⁷ Except for the conspicuous absence of language about regarding persons as means, this passage is similar to the following earlier, well-known passage in §29 of *Theory*: [#29] "to regard persons as [mere] means is to be prepared to impose on those already less favored still lower prospects of life for the sake of the higher expectations of others" (p. 157 [rev. ed.]). I think the elimination of talk of means was justified, since the disadvantaging of the less advantaged might be a merely foreseen byproduct of, rather than an instrumental means of securing, the greater advantage of the more advantaged. The fact, for example, that the net positive effect of a cut in taxation of capital gains redounds to the more advantaged to a greater extent than it redounds to the less advantaged does not necessarily imply that such a tax cut treats the latter as means.

efficient point on the OP curve would be unreasonably demanding. The objection can be made in the form of the following rhetorical question: How can you justify any alternative to the difference principle (D) involving another efficient point on the OP curve, given that, in comparison with D, those in the MAG, who would already be better off than those in the LAG, would be better off still, to the detriment of the LAG, who would not even reach the (unimproved) level of the MAG under the difference principle?³⁸ Put more simply: how can one justify benefitting the MAG who would already be better off than the LAG under the difference principle, in order to make them better off still, to the detriment of the LAG?³⁹

[#31] The rhetorical question appeals to the moral relevance of the following two considerations: **(1)** the interpersonally comparative consideration that members of a group are better or worse off than members of another group in the same society and **(2)** the counterfactually comparative consideration that members of a group are better or worse off than they would have been otherwise.

In the light of the objection via rhetorical question to any alternative to the difference principle, we are now in a position to diagnose where the argument in the previous section for choice of the principle of restricted utility from behind the veil goes wrong. Recall that, according to that argument, the choice of B^{RU} over D is in each person's rational self-interest because B^{RU} maximizes each person's expected utility from behind the veil, and it does so in a manner that insures against the downside risk of a dire outcome, on account of the suitable social minimum.

Insofar as this argument appeals to the fact that B^{RU} maximizes everyone's expected utility from behind the veil, it is indifferent to the presence or absence of complaints that members of any one group are worse off than others in society or worse off than they would have been otherwise.⁴⁰

³⁸ These alternatives to the difference principle encompass all points on the downward-sloping side of the highest OP curve. Inefficient points on or below the upward-sloping side of the OP curve would not make the MAG better off. These alternatives would be ruled out on grounds of their inefficiency.

³⁹ Cf. Otsuka and Voorhoeve, "Why it Matters that Some are Worse Off than Others: An Argument against the Priority View," *Philosophy & Public Affairs* 37 (2009): 171–99, at pp. 183–4.

⁴⁰ The insensitivity of expected utility maximization to inequality between people per se is well known. Its insensitivity to whether people are less well off than they could have been is less familiar. For a simple illustration of this latter insensitivity, consider the following "escalator case":

Sensitivity to the fact that such maximization is constrained by a guarantee that nobody's level of income and wealth fall below a suitable social minimum is also other than a sensitivity to the presence or absence of such complaints. Hence the veil of ignorance argument for restricted utility is unsound because it is insensitive to morally relevant considerations that Rawls invokes in making his reciprocity-based case for the difference principle. According to these considerations, any alternative to the difference principle would impose unreasonable demands on the least advantaged, in comparison with the most advantaged.

[#32] This insensitivity of veil of ignorance reasoning can be brought out by a consideration of Nagel's two child case with a contrasting intrapersonal version involving a single child. These contrasting cases will illustrate how a pairwise comparisons of competing claims along lines of Nagel and Scanlon diverges from impartial choice from behind the veil.

	Nagel's two-child case	
	<u>Healthy child</u>	<u>Disabled child</u>
Move to city	20	10
Move to suburb	25	6

In Nagel's two-child case depicted above, one must choose to benefit one or another of two individuals whose claims come into competition with one another.⁴¹ A move to the city would benefit one's disabled child relative to a move to the suburb, and vice versa for one's healthy child. Here one would maximize overall utility if one chose the suburb where the healthy child will flourish. Yet a pairwise comparison of the competing claims of the two children justifies a

Imagine a 100-step escalator with equally small increments between each step. Each step will be occupied by a different individual. The height of each step above the ground represents that person's absolute level of lifetime well-being. Suppose that there are only two possible outcomes, and one must choose which of them to bring about: D1 in which Persons 1-100 occupy steps 1-100 respectively, and D2 in which each person in D1 occupies one step lower, except Person 1, who occupies the very top step. (Otsuka, "Prioritarianism, Population Ethics, and Competing Claims," in J. McMahan, et al., eds., *Ethics and Existence: The Legacy of Derek Parfit* (Oxford University Press, 2022), p. 544)

Insofar as the maximization of any person's expected utility from behind a veil is concerned, it is a matter of indifference whether one chooses D1 or D2. If, however, one seeks to minimize the complaint that one is less well-off than one could have been of the person with the largest such complaint, one has decisive reason to choose D2. This is because "one person (Person 1) stands to lose an enormous amount in D1 relative to D2, whereas nobody loses much at all in D2 in comparison with D1" (ibid., p. 545).

⁴¹ See Nagel, "Equality," in his *Mortal Questions* (Cambridge University Press, 1979).

choice of the city that benefits the disabled child, since a person's lower absolute level of welfare justifies greater priority weighting of gains or losses in comparison with the gains or losses of someone else at a higher level.⁴²

[#33] In the contrasting intrapersonal one-child version depicted below of Nagel's two child case, the one child has a 50-50% chance of being disabled or able-bodied. Here one should not give priority weighting to gains or losses from a lower level of absolute welfare. Rather, one should simply maximize the one child's expected utility by choosing the suburb for him, rather than giving priority to improving his fate if he turns out disabled, by choosing the city for him.⁴³

One-child variant of Nagel's two-child case		
	50% chance	50% chance
	<u>Child if healthy</u>	<u>Child if disabled</u>
Move to city	20	10
Move to suburb	25	6

- Expected utility of city is $(0.5 \cdot 20) + (0.5 \cdot 10) = 15$
- Expected utility of suburb is $(0.5 \cdot 25) + (0.5 \cdot 6) = 15.5$

[#34] Now consider the claim that principles of social justice are those that would be chosen from behind a veil of ignorance on the assumption that one is equally likely to be any member of society.⁴⁴ This claim transforms Nagel's two-child case into the one-child variant under discussion. Rather than comparing "pairwise the positions occupied by actual people", as we did in our earlier analysis of the two-child case which yielded a priority for the worse off, it instead conceives "of those positions as slots into which one person might fall".⁴⁵ In failing to treat this

⁴² In "Prioritarianism and the Separateness of Persons," *Utilitas* 24 (2012): 365–80, I argue that the case for such priority weighting of competing claims cannot be fully explained by an aversion to the badness of inequality between actual persons in society.

⁴³ See my "Prioritarianism and the Measure of Utility," *Journal of Political Philosophy*, 23 (2015): 1–22.

⁴⁴ Recall that I argue in Appendix A that Rawls lacks good grounds for his denial of this equal probability assumption.

⁴⁵ Here I am quoting from Frances Kamm's illuminating critique of Alan Gibbard's employment of veil of ignorance reasoning. See Kamm, *Rights and Their Limits: In Theory, Cases, and Pandemics* (Oxford University Press, 2022), p. 97.

one child case differently from the two-child case, it ignores the moral significance of the separateness of persons.⁴⁶

[#35] V. The problem of non-rigid designation

[#36] The rhetorical question conveys Rawls's strongest argument on offer for the difference principle. However, this apparently compelling case for the difference principle encounters serious difficulties in scenarios in which the least advantaged do not remain the same particular individuals across different possible distributions. This is what I shall call the problem of non-rigid designation, to which I turn in the remainder of this paper. In the light of these difficulties, I shall argue that a reciprocity-based defense of the difference principle, when properly understood, must be more personal in nature than the sort of reciprocity-based justification that Rawls is willing to countenance. Such a justification is sensitive – in a way that both Rawls and his utilitarian opponents are insensitive – to the genuine moral significance of the fact that particular individuals are better or worse off than they could have been. It provides a sound case for the difference principle.

[#37] To set up this problem for Rawls, I need to add some details regarding what an OP curve represents. Rawls specifies that “A given OP curve is paired with a particular scheme of cooperation: it indicates the returns to the two groups when only wages and salaries are changed”. He also observes that “there are, in general, different OP curves for different schemes of cooperation”. How to choose among such multiple OP curves? Rawls's answer: “the difference principle directs society to aim at the highest point on the OP curve of the most effectively designed scheme of cooperation”. What qualifies as the most effectively designed scheme of cooperation? “One scheme is more effective than another if its OP curve always gives a greater return to the less advantaged for any given return to the more advantaged.” (§18.2) In other words, one scheme is more effective than another if its OP curve is higher for every value along the x-axis than the other. The OP curve in Figure 1 represents the “most effectively

⁴⁶ See Otsuka and Voorhoeve, “Equality versus Priority,” in Serena Olsaretti, ed., *Oxford Handbook of Distributive Justice* (Oxford University Press, 2018).

designed scheme of cooperation” insofar as it is, in this respect, the “highest” feasible OP curve.⁴⁷

[#38] Suppose, however, that there are two “highest” OP curves corresponding to two different schemes of cooperation, each of which is just as effective as the other, and both of which are more effective than any other feasible scheme of cooperation. We can suppose that these two different schemes of cooperation are equally effective by virtue of the fact that the OP curves associated with them are identical in their shapes, height, and length. There is, however, one crucial respect in which these two schemes differ: those particular individuals who comprise the LAG under the one scheme comprise the MAG under the other scheme, and vice versa, where, following Rawls, one’s level of income and wealth constitutes the measure of how well off one is.

Suppose that one scheme of cooperation involves an economy that specializes in IT that rewards good mathematical skills. The other involves an economy that specializes in an industry that rewards high manual dexterity. In the former scheme of cooperation, those with good mathematical skills constitute the MAG and those with good manual dexterity constitute the LAG. In the latter scheme of cooperation their positions are switched: those with good manual dexterity constitute the MAG and those with good mathematical skills constitute the LAG. Suppose, as Rawls allows, that the different talents of these different individuals track their genomes.⁴⁸ Let us also suppose, for simplicity, that the LAG and the MAG are always equally large groups.

Here we have a choice between an IT economy and an industrial economy, in which different individuals would be among the LAG and the MAG under different economic arrangements, on

⁴⁷ Rawls maintains that “[a] When these [OP] curves criss-cross, the one tangent to the highest JJ line [i.e., the highest line parallel to the y-axis] is best; [b] if they touch the same JJ line, the one whose tangent is to the left of the other is best.” (§18.2, n. 32) I believe that [a] begs the question in favor of the difference principle over restricted utility by privileging schemes of cooperation with the highest D point over those whose B^{RU} point is greatest. Frances Kamm has noted that [b] implies an endorsement of a levelling down of the prospects of the MAG. I would add that [b] is also inconsistent with Rawls’s lexical formulation (about which see *Theory*, p. 72 [rev. ed.]) of the difference principle.

⁴⁸ In §21.3, Rawls says that both “variation of talents of the same kind (variation in strength and imagination, and so on)” and “the variety of talents of different kinds” might involve “differences among persons” in their “native endowments”. In §21.4, Rawls refers to the “natural fact of the distribution of endowments” where “some are by nature better endowed than others”.

the plausible assumption that the identities of these individuals track their native endowments. In fact, the people in the LAG and the MAG swap places under the two schemes.⁴⁹

[#39] This scenario involving a choice between two equally most effective schemes of cooperation (IT and Industry) poses the following problem for Rawls's defense of the difference principle. Recall that Rawls says that "D is the only point on the (highest) OP curve that meets the following reciprocity condition: those who are better off at [that] point are not better off to the detriment of those who are worse off at that point" (§36.3). The realization of the difference principle under the one scheme of cooperation is, however, "to the detriment of" the worse off in the following respect. Those particular individuals, who are members of the worse off group in the one scheme of cooperation regulated by the difference principle, would be much better off under an equally good alternative scheme of cooperation, where that scheme of cooperation is regulated by the principle of restricted utility rather than the difference principle. This is both because these particular individuals would be members of the MAG rather than the LAG, and because they would be at point B^{RU} to the right of point D of an identically-sized-and-shaped OP curve under the alternative scheme of cooperation. Hence, the difference principle would be to the detriment of these particular individuals, in comparison with restricted utility.⁵⁰ For example, those particular manually dexterous individuals who are the members of the least advantaged group in the IT economy – Mechanics at \$50k in D (IT) in the figure below – would be as well off as they could possibly be in an IT economy when it is regulated by the difference principle. These individuals would, however, be much better off in an industrial economy regulated by a principle of restricted utility, as they would be members of the most advantaged group in that case, with a level of income and wealth that places them far along the x-axis. They would be Mechanics at \$100k in B^{RU} (Industry).⁵¹

⁴⁹ This scenario is consistent with Rawlsian "fair equality of opportunity" (as defined in §13.2) for the following reason: all who have the same native endowments have the same prospects and only those with different native endowments have different prospects. What varies across the two schemes of cooperation is whether the equal prospects which correspond to a particular set of native endowments are equally high or equally low prospects.

⁵⁰ These particular individuals would also have been better off under the alternative scheme of cooperation even if it were also regulated by the difference principle, since they would have been members of the MAG at point D rather than the LAG.

⁵¹ The same holds for a comparison of D (Industry) with B^{RU} (IT), except that different particular individuals – Quants rather than Mechanics – would be at \$50k and \$100k in D (Industry) with B^{RU} (IT) respectively.

	Mechanics	Quants
D (IT)	\$50,000	\$75,000
B^{RU} (IT)	\$40,000	\$100,000
D (Industry)	\$75,000	\$50,000
B^{RU} (Industry)	\$100,000	\$40,000

Why, one might ask, does Rawls need to consider a scenario in which the difference principle and the principle of restricted utility are applied across the different schemes of cooperation represented by an IT and an industrial economy? The answer is that the case for the difference principle over a principle of restricted utility ought to be robust to such a transformation of schemes of cooperation. In particular, one should be able to justify, to those who would be the LAG under a given scheme of cooperation, the choice of that scheme over another in which they would be the MAG. Why should they be consigned to a scheme of cooperation in which they are the LAG when there is another equally effective scheme in which they are the MAG? Hence the need, in such a scenario in which there are two equally effective schemes of cooperation (two highest OP curves), to provide a justification, not just of the difference principle over a principle of restricted utility, but of one equally effective scheme over the other.⁵² If the difference principle is to be justified, it must be shown to be justifiable to each, as against the feasible alternatives, including B^{RU} plus a swapping of schemes of cooperation.

[#41] The above problem arises for Rawls on account of the “rigid designation” of the individuals who are members of the least advantaged group. In *Justice as Fairness*, however, Rawls makes clear that the term “the least advantaged” should not be read as rigidly designating the individuals who compose the membership of this group in a given scheme of cooperation:

Taking these cooperative schemes as possible social worlds (let’s say) over which the names of individuals refer to (rigidly designate) the same individuals in each possible

⁵² [#40] One might object that it’s unrealistic to assume that the two curves are exactly the same shape. It is, however, realistic to suppose that, given all we would know in the actual society in which we make the choice regarding the one specialization or the other, we would have no reason to assume that the curves differ in shape even though we realize that, in actual fact, they won’t be exactly the same shape.

(social) world, the term “the least advantaged” is not a rigid designator (to use Saul Kripke’s term, see *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1972)). Rather, the worst off under any scheme of cooperation are simply the individuals who are worst off under that particular scheme. They may not be those worst off in another. (§17.3 n. 26)

Insofar as application of the difference principle is concerned, “the least advantaged” are simply those individuals who have the least income and wealth under different possible distributions. Those who have the least income and wealth needn’t be the same individuals under different schemes of cooperation.

[#42] Rawls objects as follows to a justification of the difference principle to particular rigidly designated individuals who are members of the least well off group under a given scheme of cooperation:

the difference principle does not appeal to the self-interest of those particular persons or groups identifiable by their proper names who are in fact the least advantaged under existing arrangements, rather, it is a principle of justice.⁵³ In ideal theory, the only defense of inequalities in the basic structure is that they make the worst off (whoever they may be...) better off than the worst off (whoever they may be) under any alternative (practicable) scheme consistent with all the requirements of the two principles of justice. In this way, the difference principle expresses, as any principle of political justice must, a concern for all members of society. (§19.5)⁵³

[#43] An appeal to the fact that the particular rigidly designated worst off individuals are less well off than they would have been under a different arrangement need not, however, simply involve an appeal to self-interest, which is of no moral significance to matters of justice. It needn’t only be a matter of partial self-interest, rather than impartial moral concern. One can

⁵³ Here Rawls is responding to a challenge posed by an example of John Broome’s in which, as in the case of my example involving an industrial versus an IT economy, the fates of the rigidly designated and the non-rigidly designated worst off diverge. In Appendix B of this paper [which is available [via this link](#)], I discuss further attempts of Rawls to fend off the challenge of Broome’s example. There I explain how these attempts fall short and also why my example poses a greater challenge to Rawls than Broome’s does.

recognize the moral significance of a particular individual's being worse off than she would have been, even from an impartial perspective. When proper names are a proxy for such considerations that track the fates of particular individuals across different arrangements, attention to proper names is not attention to a factor that is irrelevant from a moral point of view. That it is someone named 'Sam' rather than 'Jamie' should not make a moral difference. But whether someone, tracked by proper name, is worse off than that same person would have been otherwise is morally relevant.

[#44] When, following Rawls, one eschews rigid designation, problems arise for his justification of the difference principle. Recall that Rawls maintains that "the difference principle expresses the idea that ... the more advantaged are not to be better off at any point *to the detriment of* the less well off" (§36.4, my emphasis added). The worse off group can be made better or worse off in non-person tracking terms that do not rigidly designate: i.e., those who are worse off can be made better or worse off than those who are worse off would have been, even if the membership of the worse off group consists of entirely different individuals. But I would maintain that something can be *to the detriment of* the less advantaged, or that it can be contrary to their interests, only if the particular individuals who constitute this group are made worse off in person-tracking terms that rigidly designate particular individuals. Similarly, something can be to the benefit of the more advantaged only if its members are made better off in person-tracking terms that rigidly designate.

I'm not advancing the more general claim that something can never be to the detriment or benefit of a group of individuals if its rigidly designated members are not made better or worse off. Perhaps something can be to the detriment or the benefit of some corporate entities even in the absence of any detriments or benefits to any rigidly designated members. Under certain conditions of class consciousness and identification of individuals with the interests of the collective, 'the proletariat' might count as among them. But the less or more advantaged, as picked out simply by their relative level of income and wealth, is not among them.⁵⁴

⁵⁴ Even if the least advantaged identified as members of such a proletariat, I'm doubtful that such identification with the interests of the working class would track its fate across a different scheme of cooperation in my example under discussion, in which such identification would need to counterfactually extend from the manually dexterous to the entirely disjoint and very different collective of the mathematically talented who occupy the managerial class in the actual world.

[#45] The following joke of Zsa Zsa Gabor's drives this point home:

Zsa Zsa. ...I am constantly finding new ways to do good for people.

Interviewer. Like what?

Zsa Zsa. I have found a way of keeping my husband young and healthy, almost forever.

Interviewer. Eternal youth . . . that is quite a discovery! How do you do it?

Zsa Zsa. I get a new one every five years!⁵⁵

Zsa Zsa's non-rigidly designated "husband" remains eternally young and healthy down through the decades insofar as successive occupants of this role are young and healthy. Yet no actual individual experiences such a boon of eternal youth and health. That's why it strikes us as absurd for Zsa Zsa to maintain that she is doing good for anyone by remarrying every five years. We should say something similar about increases or decreases in the level of income and wealth of the least advantaged group that do not involve any benefits or detriments to actual rigidly designated individuals.

It would be a mistake to maintain that any alternative to the difference principle would be to the detriment of the least well off simply by virtue of the fact that this would make the non-rigidly designated least well off less well off. To show why, let us consider the analogous claim that Zsa Zsa's remaining married to the same person for life, rather than remarrying every five years, would be to the detriment of her husband because it would make him much older than he would have been otherwise. It is true that Zsa Zsa's non-rigidly designated husband grows much older in a scenario in which she remains married for life, whereas her non-rigidly designated husband remains young in the alternative scenario in which she remarries a younger man every five years. There is, however, no individual to whom it makes sense to attribute a detriment of growing older than he would otherwise have been. This is because a non-rigidly designated husband does not experience any benefits or detriments in scenarios in which the term "husband" picks out different rigidly designated individuals across different possible worlds. There is no person, much less the same person, who remains young in this scenario. The young person-stages belong to different people, according to our criterion of identity over time. To maintain otherwise is to

⁵⁵ As quoted by Caspar Hare, in his "Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?" *Ethics* 117 (2007): 498–523, at 514.

license the blending together of different rigidly designated individuals across different possible worlds as if they belonged to a single life. This would ignore the separateness of persons.

Growing old would have been to the detriment of her husband in a scenario in which Zsa Zsa remains married for life, only if the particular rigidly designated individual who ages would have been younger if Zsa Zsa had remarried every five years. But he would not have. He would have aged at exactly the same rate. Hence, at least insofar as chronological rate of aging is concerned, Zsa Zsa's remaining married to her husband would not be to her husband's detriment. This is not akin to a scenario in which one is deprived of waters from a fountain of youth from which one was about to drink.

[#46] When one restricts oneself to claims that do not track the fates of particular rigidly designated individuals, it becomes difficult to justify the moral significance that Rawls attributes to moves that are Pareto improvements in comparison with those that involve conflicts of interest. In his discussion of Figure 1, Rawls draws a contrast between moves that are to the mutual advantage of the MAG and the LAG on the one hand, and those that involve trade-offs between the MAG and the LAG on the other hand. Rawls maintains that one should move from the origin to point D, since "everyone benefits" from such moves. They are strong Pareto improvements. He also maintains that one should stop at point D, since one would then be entering the "conflict segment" involving trade-offs. The distinction between Pareto-improvements and trade-offs appeals to the fact that nobody is made worse off to secure the overall greater good in the former case, whereas the interests of some are sacrificed in the latter case.⁵⁶ But, as I shall show below, Rawls loses this distinction, with his non-rigid-designation of groups in a manner that fails to track the fates of particular rigidly designated individuals.

⁵⁶ Recall the following contrast that Rawls draws between the reciprocity-based case for the difference principles versus the case grounded in sympathy for the principle of restricted utility:

For as a principle of reciprocity, the difference principle rests on our disposition to respond in kind to what others do for (or to) us; while the [restricted] utility principle puts more weight on what is a considerably weaker disposition, that of sympathy, or better, our capacity for identification with the interests and concerns of others. (§38.2)

The disposition to respond in kind is satisfied by mutually advantageous gains, whereas sympathy or identification with the interests and concerns of others is required in order to be moved by the fact that the gains to others are greater than the losses to oneself when the interests of individuals are in conflict.

[#47] Recall that Rawls maintains that “the worst off under any scheme of cooperation are simply the individuals who are worst off under that particular scheme. They may not be those worst off in another.” For the purposes of applying the difference principle, the worst off are simply those who have the least income and wealth.

In the absence of the rigid designation of the particular individuals who occupy this and other groups that are ranked by level of income and wealth, one is unable to register the morally significant difference between cases such as the following two, the first of which involves no trade-offs and the second of which involves trade-offs:

Case One	D1		D2	
Ann>>	MAG	\$75k Ann	MAG	\$100k Ann
Ben>>	LAG	\$40k Ben	LAG	\$50k Ben
No trade-off				

Case Two	D1		D2	
Ann>>	MAG	\$75k Ann	LAG	\$50k Ann
Ben>>	LAG	\$40k Ben	MAG	\$100k Ben
Trade-off				

In these tables, the fates of rigidly designated individuals are tracked by the rows, the columns of which are occupied by rigidly designated individuals as identified by proper name. In Case One, but not Case Two, each such individual is better off in D2 than they are in D1. The different colors of the cells track the non-rigidly designated LAG and MAG, ranked by level of income. There is no difference between Case One and Case Two when groups are specified only in this manner. Hence, the moral significance of presence or absence of trade-offs is lost.

[#48] Now consider the following contrast between versions of the illustrated distributions under the difference principle and restricted utility where rigidly designated individuals switch ranks and where they do not switch ranks.

Case 1	B ^{RU}	D
Amy	\$100k	\$75k
Bob	\$40k	\$50k

Case 2	B ^{RU}	D
Amy	\$100k	\$50k
Bob	\$40k	\$75k

The choice of one outcome over the other is *non-rank-switching* in Case 1, whereas it is rank-switching in Case 2, where a *rank-switching* choice is one that makes a difference to which rigidly designated individuals are better or worse off than other rigidly designated individuals.

In Case 1, Bob could object to Amy's insistence on B^{RU} over D by deploying the rhetorical question: How can you justify B^{RU} over D when you would remain better off than me in D, while B^{RU} would make you better off still, relative to D, and to my detriment? A comparable complaint against D is unavailable in Case 2, since Amy would not remain better off than Bob in his preferred outcome D. Rather, there would be a switch in their ranking, and Amy would be worse off than Bob in D. The rhetorical question serves to condemn only refusal to engage in the non-rank switching change. It doesn't also condemn refusal to engage in the rank-switching change. The presence or absence of the complaint embodied by the rhetorical question is a matter of how rigidly designated individuals fare relative to one another. Case 1 is indistinguishable from Case 2 if one doesn't track the fates of rigidly designated individuals. But the rhetorical question reveals that it matters how particular rigidly designated individuals would fare in comparison with others who are better off. This moral fact is lost if, as Rawls does, one non-rigidly designates the worst off.

[#49] VI. The best theoretical account of the underlying motivation of contractualism implies rigid designation

[#50] In his discussion of Singer's example of saving the child in the pond, Scanlon draws a contrast between a motivation to save the drowning child because this would maximize the sum total of happiness and saving that child because one could not justify a failure to do so to that child, when the cost to oneself of doing so is so small, by comparison with what's at stake for the

child. Utilitarianism involves an “abstract” concern regarding “changes in aggregate well-being, however these may be composed”.⁵⁷ According to contractualism, by contrast:

the source of motivation that is directly triggered by the belief that an action is wrong is the desire to be able to justify one’s actions to others on grounds they could not reasonably reject. I find this an extremely plausible account of moral motivation - a better account of at least my moral experience than the natural utilitarian alternative.⁵⁸

Scanlon’s account of why he couldn’t justify his failing to aid to the individuals who would die appears to involve an inability to justify oneself to particular, rigidly designated individuals, rather than more abstract non-rigidly designated placeholders such as ‘those with the lowest income and wealth, whomever they might turn out to be’, or ‘my husband, whomever he might turn out to be.’

Suppose that Zsa Zsa were to claim that she could not justify a failure to supply waters from the fountain of youth to her husband, since she is so wealthy that it would be little sacrifice to her to purchase such waters. This claim has force. Now consider the claim that she could not justify remaining married to her husband, rather than remarrying a younger man every five years, on grounds that this would make her husband older than he would otherwise have been. This claim comes across as absurd. A joke. To see why, let us consider the following:

[#51] For Scanlon, morality involves justification to beings with a perspective: “morality applies to a being if the notion of justification to a being of that kind makes sense.”⁵⁹ He also writes:

[A] minimum requirement for this notion [of justification to a being to gain a foothold] is that the being constitute a point of view; that is, that *there be such a thing as what it is like to be that being*, such a thing as what the world seems like to it. Without this, we do

⁵⁷ “Contractualism and Utilitarianism,” in *Utilitarianism and Beyond*, eds. Amartya Sen and Bernard Williams (Cambridge University Press, 1982), p. 115

⁵⁸ “Contractualism and Utilitarianism,” p. 116. See also Scanlon’s more recent discussion of Singer’s example in “Contractualism and Justification,” in his *Morality and Responsibility* (Polity, 2025), pp. 40-41, 52.

⁵⁹ “Contractualism and Utilitarianism,” p. 113.

not stand in a relation to the being that makes even hypothetical justification *to it* appropriate.⁶⁰

This explains the moral significance of consciousness: “If a being can feel pain, then it constitutes a centre of consciousness to which justification can be addressed.”⁶¹

Groups of human beings aren’t conscious beings with perspectives.⁶² Although Scanlon doesn’t explicitly draw this connection, I think this fact can help explain why Scanlon restricts his contractualism to the claims of individuals rather than groups.

[#52] Perspective is also crucial for Nagel:

the impersonal concern that results is fragmented: it includes a separate concern for each person, and it is realized by looking at the world from each person’s point of view separately and individually, rather than by looking at the world from a single comprehensive point of view. Imaginatively one must split into all the people in the world, rather than turn oneself into a conglomeration of them.⁶³

Moreover, contractualism involves:

a point of view which abstracts from who we are, but which appreciates fully and takes to heart the value of every person’s life and welfare. We put ourselves in each person’s shoes and take as our preliminary guide to the value we assign to what happens to him the value which it has from his point of view.⁶⁴

⁶⁰ “Contractualism and Utilitarianism,” p. 114.

⁶¹ “Contractualism and Utilitarianism,” p. 114. This is perhaps more plausible as a necessary but not a sufficient condition of the possibility of justification.

⁶² In the example above, Zsa Zsa’s non-rigidly designated husband consists of the following group: a sequence of young person-stages, each of which occupies the role of Zsa Zsa’s husband for a five-year period. Each of these person-stages is part of a person with a perspective. But, from the perspective of each person, what matters is their fate over a lifetime that does not coincide for more than a five-period period with the history of Zsa Zsa’s non-rigidly designated “husband”.

⁶³ “Equality,” p. 127.

⁶⁴ *Equality and Partiality* (Oxford University Press, 1991), pp. 64-5.

[#53] Morality involves justification to beings with a perspective. But a non-rigidly designated position is not a being with a perspective. E.g., Zsa Zsa's non-rigidly designated 'husband' is not a being with a perspective.

[#54] VII. How to justify Rawls's non-rigidly designated difference principle to rigidly designated individuals

I would like, at this point, to acknowledge that Rawls has the following compelling reason to adopt the non-rigid designation of the LAG in spelling out the meaning of the difference principle. In making the least advantaged as well off as possible, we shouldn't pick out those rigidly-designated individuals who are least well off under a given distribution and make *these particular individuals* as well off as we can, even if this has the upshot that others become much less well off than they. In our scenario involving two possible schemes of cooperation (IT and Industry), such a reading of the difference principle might imply that the rigidly designated least advantaged individuals under the one scheme of cooperation be transformed into the most advantaged by adopting the other scheme of cooperation and then raising them to the level of the MAG at the feudal point F which lies far to the right of the x-axis. [>#55<] This absurdity is avoided by adoption of a non-rigid designation of the least advantaged in interpretation of the directive to make them as well off as possible. So my complaint is not with Rawls's adoption of the non-rigid designation in his *spelling out* of his difference principle. Rather, my complaint is with Rawls's *defense* of the difference principle by means of an attempt to justify it to non-rigidly designated individuals.

Rather than trying to offer a justification to non-rigidly designated individuals (which I have shown to be problematic), one might offer the following justification of Rawls's non-rigidly designated difference principle to all particular (rigidly designated) individuals, including the rigidly designated worst off. One might be able to do so, even when, as in our IT versus industrial economy case, the rigidly designated worst off would be better off under an alternative to the difference principle.

[#55] We could say the following to the rigidly designated worst off individual under the difference principle (who is at \$50,000 in IT regulated by the difference principle): although you would be better off otherwise (you would be at \$100,000 in Industry regulated by restricted

utility), the rigidly-designated worst off individual under this alternative to the difference principle would have the complaint that they would have less income and wealth (\$40,000 in Industry regulated by restricted utility) than the worst off individual would have under the difference principle. In other words, even when offering a justification only to rigidly designated individuals, one might still be able to justify the difference principle, on grounds that any alternative to it would result in a particular individual's being worse off than any particular individual would be under the difference principle.⁶⁵

	Mechanics	Quants
D (IT)	\$50,000	\$75,000
B ^{RU} (IT)	\$40,000	\$100,000
D (Industry)	\$75,000	\$50,000
B ^{RU} (Industry)	\$100,000	\$40,000

[#56] Consider what Scanlon calls the ‘Complaint Model’: “On this interpretation of contractualism, a person’s complaint against a principle must have to do with its effects on him or her, and someone can reasonably reject a principle if there is some alternative to which no other person has a complaint that is as strong.”⁶⁶

Cf. Nagel, “Where there is conflict of interests, no result can be completely acceptable to everyone. But it is possible to assess each result from each point of view to try to find the one that is least unacceptable to the person to whom it is most unacceptable.”⁶⁷

⁶⁵ Though he doesn’t characterize it as a justification to rigidly designated individuals, Rawls offers such a justification of the difference principle in response to Broome’s example in which the fate of the non-rigidly designated least advantaged doesn’t track the fates of rigidly designated individuals. Rawls maintains that we should choose the difference principle here on grounds that “the worst off would be even worse off” under the alternative. (§19.4)

⁶⁶ *What We Owe to Each Other* (Harvard University Press, 1998), p. 229.

⁶⁷ “Equality,” p. 123.

[#57] Scanlon maintains that “Under contractualism, when we consider a principle our attention is naturally directed first to those who would do worst under it. This is because if anyone has reasonable grounds for objecting to the principle it is likely to be them.”⁶⁸

Cf. Nagel: “The preferred alternative is ... the least unacceptable, considered from each person’s point of view separately. A radically egalitarian policy of giving absolute priority to the worst off, regardless of numbers, would result from always choosing the least unacceptable alternative, in this sense.”⁶⁹

One who adopts a principle of minimax complaint might therefore argue that nobody could reasonably reject the difference principle since it uniquely minimizes the maximum complaint, when this is measured by the magnitude of any shortfall in the level of income and wealth of the rigidly designated worst off person, relative to the level of the income and wealth of the rigidly designated worst off person under the alternatives. There would be no such shortfall if the difference principle is adopted. By contrast, any alternative to the difference principle could be reasonably rejected, since it would be at the expense of rigidly designated individuals being worse off than those rigidly designated individuals who are worst off under the difference principle.

[#58] There is, however, the following problem with such a minimax complaint-based attempt to justify the difference principle to rigidly designated individuals: It zeros in how badly off the worst off person is under one distribution relative to how badly off the worst off person is under other distributions as the only component relevant to the magnitude of complaints. But this isn’t the only thing that’s relevant. As Scanlon and Nagel point out, other factors are also relevant.⁷⁰ How much those rigidly designated individuals who are members of the MAG would gain if the position of the LAG isn’t maximized, is relevant. How well off those rigidly designated individuals who are members of the LAG would be under alternatives to the difference principle is also relevant. How well off individuals are in comparison with others in their society under a given principle of distribution is also relevant to the magnitude of their complaints. Does a case

⁶⁸ “Contractualism and Utilitarianism,” p. 123.

⁶⁹ “Equality,” p. 123.

⁷⁰ The complaint model that Scanlon considers in “Contractualism and Utilitarianism” is sensitive to such considerations. What Scanlon calls “welfarist contractualism” (which he doesn’t endorse) is similarly sensitive. See *What We Owe*, pp. 217-18, 242-3.

for the difference principle remain when all factors relevant to complaints are considered? I now turn to this question.

[#59] Let me provide an illustration of these different factors that comprise the complaints of individuals, in a pairwise comparison of the complaints of rigidly designated individuals in the case of the industrial versus the IT economy.

	Mechanics	Quants
D (IT)	\$50,000	\$75,000
B ^{RU} (IT)	\$40,000	\$100,000
D (Industry)	\$75,000	\$50,000
B ^{RU} (Industry)	\$100,000	\$40,000

As discussed above, the IT versus industrial example demonstrates that, when the LAG are rigidly designated, the alternative to the difference principle is actually to the detriment of the LAG. These rigidly designated individuals therefore have a complaint against the difference principle, as compared with the alternative of restricted utility. The green highlighting tracks their fate under the difference principle as compared to this alternative. Their complaint against the difference principle is that they would be both worse off under the difference principle than they would have been under restricted utility (\$50k rather than \$100k – see 1st column), and that they would be worse off than others under the difference principle (\$50k rather than \$75k – see first row).

Recall the above discussion that what's morally relevant are the complaints of rigidly designated rather than non-rigidly designated individuals. In arguing this, I drew heavily on the case of Zsa Zsa Gabor's husband.

If what matters are the complaints of rigidly designated individuals, and since the difference principle is actually *to the detriment of* the rigidly designated least advantaged, relative to restricted utility, in our IT versus industrial example, does it follow that we should embrace restricted utility over the difference principle in this example?

Not necessarily! We also need to see what sort of complaint, if any, the rigidly designated least advantaged individuals would have under restricted utility, in comparison with the difference principle. We can see that the rigidly designated least advantaged individuals would also have a complaint against restricted utility, relative to the difference principle. In Industry regulated by B^{RU} (rather than IT regulated by D) the least advantaged Quants can complain both that they're worse off than they would have been, at \$40k versus \$75k, and that they're worse off than others, \$40k versus \$100k. They can also complain that they're worst off than the worst off would be otherwise: \$40k, rather than \$50k.⁷¹

Who has the greater complaint?

[#60] If the only thing relevant to the measure of the complaint of the rigidly designated worst off were the lowness of their income and wealth, relative to the level of income and wealth of the rigidly designated worst off under an alternative distribution, then we could conclude that the rigidly designed least advantaged Quants in Industry regulated by B^{RU} have the greatest complaint, since their income is at \$40k, whereas the income of the rigidly designated least advantaged Mechanics in IT regulated by D are at \$50k.

But, as I indicated earlier, comparisons of level of the income and wealth least advantaged across different alternatives is not all that's relevant to one's level of complaint. It's also relevant how much better off one would have been otherwise: i.e., how much of a detriment one suffers, as compared with the alternative. The better off one would have been otherwise, the greater one's complaint.

Insofar as this element of a complaint is concerned, the rigidly designated least advantaged Mechanics in IT regulated by D have a greater complaint against D than the rigidly designated least advantaged Quants in Industry regulated by B^{RU} have against B^{RU} . This is because the extent to which the Mechanics in IT regulated by D are worse off than they would have been otherwise is greater than the extent to which the Quants in Industry regulated by B^{RU} are worse off than they would have been otherwise. This is greater, when measured both in terms of

⁷¹ This low level of \$40k is, however, mitigated by the fact that it is above a suitable social minimum.

interval and ratio. The Mechanics in IT regulated by D are \$50k worse off than – i.e., half as well off as – they would have been under the alternative available to them. By contrast, the Quants in Industry regulated by B^{RU} are \$35k worse off than – and 53% as well off as – they would have been under the alternative available to them.

We also need to assess complaints against the level of interpersonal inequality within a given scheme of cooperation. Here the rigidly designated least advantaged Quants in Industry regulated by B^{RU} have the stronger complaint than the least advantaged Mechanics in IT regulated by D: a \$60k gap and 40% as well off, v. a \$25k gap and 67% as well off as the MAG in the same scheme.

Typically, we would also have to assess the strength of the complaints of the rigidly designated better off and compare them with the strength of the complaints of the rigidly designated worse off. I believe, however, that such comparisons are already captured in the above, given the fact that each rigidly designated member of the LAG is a member of the MAG under the alternative.

Who has the greater complaint, all things considered? The Quants who would be least advantaged in Industry regulated by B^{RU} or the Mechanics who would be least advantaged in IT regulated by D? Since the different complaints pull in different directions, it ultimately depends on the weight one places on lowness of income and wealth of those with the least in one distribution relative to another distribution, detriment to individuals in comparison with how they would fare in the alternative, and societal inequality. On the one hand, both level of inequality within a scheme of cooperation and lowness of the level of the least well off in a comparison across schemes tell in favor of the D and against B^{RU}. On the other hand, the complaint that one is worse off than one would have been in the alternative tells against D and in favor of B^{RU}.⁷² In considering both number and strength of complaints, I think the balance tips in favor of D over B^{RU}. It would be unreasonable, here, to place so much weight on how well off rigidly designated individuals would have been in the alternative that this outweighs the combination of the other two complaints.

⁷² Under B^{RU}, unlike an unrestricted version of a principle of utility, nobody will ever have the further complaint that their level of income and wealth falls below a suitable social minimum. Since, however, the least well off under D are better off than the least well off under B^{RU}, nobody in D will have such further complaint either. Therefore, absence of such a complaint does not tell in favor of either principle over the other one.

It is on account of facts regarding the shape of the OP curve that the balance of complaints tells in favor of the difference principle. If the OP curve had been flatter to the left and right of point D and more elongated along the x-axis, then the balance of complaints might tell in favor of B^{RU} rather than D. This is because individuals would be much better off in the restricted utilitarian alternative to the difference principle and the least well off individual under restricted utility would not be much less well off than the least well off individual under the difference principle. Rawls contends, and I shall assume throughout this discussion, that the OP curve would not have such a flat shape when the prior principles of liberty and fair equality of opportunity plus the opportunities for transfers from the MAG to the LAG are fully taken into account.⁷³

On account of this contingency of the case for it on the shape of the OP curve, the difference principle is not a fundamental principle of social justice. Rather, it is derived from a principle of minimax complaint, which is more fundamental insofar as it is more directly derived from the underlying principles of Scanlonian/Nagelian contractualism.

[#61] To take stock of where we now are: I argued above that what's morally relevant are the complaints of rigidly designated rather than non-rigidly designated individuals. I then showed how the complaints of rigidly designated individuals should be assessed in the case of an industrial versus an IT economy. The upshot of such comparison is that, even when one engages in a sound method of justification of distributive principles to rigidly designated individuals, the case for the difference principle still goes through.

[#62] I would now like to draw attention to the following fact. When, as is usually implicitly assumed in discussions of the difference principle, the same rigidly designated individuals constitute the LAG across the different possible options, the rigidly designated members of the LAG have no complaint against D. All elements of their complaint are against the alternative of B^{RU} : they would be at a lower level of \$40k under B^{RU} than their level of \$50k under D and would therefore suffer a detriment under B^{RU} relative to D. They would also suffer greater societal inequality under B^{RU} than under D.

⁷³ See §§19.1-19.2.

	Mechanics	Quants
D (IT)	\$50,000	\$75,000
B^{RU} (IT)	\$40,000	\$100,000
D (Industry)	\$75,000	\$50,000
B^{RU} (Industry)	\$100,000	\$40,000

It doesn't follow from this fact that the rigidly designated members of the LAG have the greatest complaint, which is against B^{RU} . We also need to assess the magnitude of the complaint of the rigidly designated members of the MAG, which is against D. However, the 'rhetorical question' combined with the aforementioned facts regarding the non-flatness of the OP curve demonstrate that those in the LAG have a greater complaint against B^{RU} than those in the MAG have against D.

When the fates of rigidly designated individuals perfectly track the fate of the non-rigidly designated LAG, the case for the difference principle is therefore more decisive than it is in my scenario involving two different schemes of cooperation in which the fates of rigidly designated individuals come apart from the fate of the non-rigidly designated LAG. Since it registers complaints only of non-rigidly designated groups, Rawls's own justification for the difference principle is insensitive to this difference. It needs to be rejected on that account in favor of the method of justification that I have proposed.

[#63] In drawing this discussion to a conclusion, I shall turn to an examination of the case for the difference principle in non-identity cases. This will reveal a further shortcoming with Rawls's justification of the difference principle on account of its restriction to the fates of non-rigidly designated individuals.

I shall begin by noting that non-identity cases are not of merely theoretical interest. Choices between different principles of distribution would give rise to actual non-identity cases, for the following reason.

On account of the split-second contingency of conception, the choice of the difference principle versus restricted utility would make a difference to the identity of nearly everyone who is conceived from the point of implementation of the one principle of distribution over the other. Therefore, within 20 years, those entering the workforce in a society regulated by the one principle of distribution would not have existed had the other principle of distribution been enacted instead. In about 80 years, the identity of just about everyone in the workforce will be contingent on choice of the difference principle versus restricted utility.

Suppose, now, that the principle of restricted utility is enacted today and we ask whether such a scheme of cooperation can be justified over the long term, understood as 80 years from now and beyond. Would such a scheme regulated by restricted utility be ‘to the detriment of’ the LAG in the long term, in comparison with one regulated by the difference principle? To be sure, the non-rigidly designated LAG would be worse off than the LAG would otherwise be, since those earning the lowest income earn \$40k rather than \$50k. But would this really be ‘to the detriment of’ the least advantaged? It’s hard to make sense of the claim that it would be, for the following reason: under restricted utility, the rigidly designated members of the LAG would have lives worth living, since they would be earning at least a suitable social minimum. And, had the difference principle been enacted in the past instead of restricted utility, they would not have existed.⁷⁴ I believe that it follows that restricted utility is better than the difference principle for the rigidly designated members of the LAG who exist under restricted utility, since life at \$40k is better for them than non-existence. As Arrhenius and Rabinowicz have argued, existence might be better for someone than non-existence, even if it is not the case that non-existence would have been worse for them if they had not existed.⁷⁵ At the very least, it is uncontroversially the case that the rigidly designated members of the LAG under restricted utility would not have been better off under the difference principle, since they would not have existed.

⁷⁴ One might ask: Could we not improve the lot of the rigidly designated LAG from \$40k to \$50k in 80 years’ time by effecting a transition of that scheme of cooperation at that point from one regulated by B^{RU} to one regulated by D? In reply, I note: (i) It is plausible to suppose that, on account of the transition cost of moving from B^{RU} to D after having first implemented B^{RU} , the D point one arrives at would not be on the highest feasible OP curve. Hence it would be ruled out. See earlier discussion in [#37]. (How do you get to D? I wouldn’t start from here, where “here” is B^{RU} .) (ii) A long-term non-identity problem will simply resurface even if there is a contemporaneous justification of the transition from B^{RU} to D.

⁷⁵ Gustaf Arrhenius and Wlodek Rabinowicz, “The value of existence”, in *Oxford Handbook of Value Theory*, eds. Iwao Hirose and Jonas Olson (Oxford University Press, 2015), pp. 424–444. By the same reasoning, it’s also the case that the difference principle is better than RU for the rigidly designated MAG who exist under the difference principle. This is because the MAG have lives worth living under the difference principle. And they would not have existed under RU.

	M	M*	Q	Q*
D (IT)	\$50k		\$75k	
^{RU} B (Industry)		\$100k		\$40k

Under restricted utility, the rigidly designated members of the LAG might object to the fact that they are worse off than others. But since their lives are well worth living – as they are above the social minimum – and they would not have existed otherwise, this fact of inequality does not give rise to a complaint against restricted utility, relative to the difference principle, all things considered.

Here now is the problem for Rawls's justification of the difference principle, which this discussion of a non-identity case reveals. Rawls is unable to distinguish the strength of the case for the difference principle in a non-identity scenario in comparison with a scenario in which the non-rigidly designated LAG always consists of the same rigidly designated individuals. The case is just as strong, when, as Rawls does, one restricts oneself to a consideration of detriments and benefits to non-rigidly designated groups. But surely the case for the difference principle is much stronger when any alternative to it is to the detriment of rigidly designated individuals than it is in such a non-identity case.