View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by PhilPapers

The Pareto Argument for Inequality Revisited[1]

A. R. J. Fisher & E. F. McClennen

Abstract: *one of the more obscure arguments for Rawls' difference principle dubbed 'the Pareto argument for inequality' has been criticised by G. A. Cohen (1995, 2008) as being inconsistent. In this paper, we examine and clarify the Pareto argument in detail and argue (1) that justification for the Pareto principles derives from rational self-interest and thus the Pareto principles ought to be understood as conditions of* individual rationality, *(2) that the Pareto argument is not inconsistent, contra Cohen, and (3) that the kind of bargaining model required to arrive at the particular unequal distribution that the difference principle picks out is a model that is not based on bargaining according to one's threat advantage.*

The Pareto argument for Inequality is one of the more obscure arguments for Rawls' difference principle found in *A Theory of Justice*. Put briefly, the argument justifies a *particular* deviation from an initial equal distribution of primary social goods that is ultimately unequal so long as everyone is made better off from this deviation. The argument itself is arguably the 'more fundamental' and 'more direct' argument for the difference principle.[2] In addition, we say, it is one of the 'strongest' arguments for the difference principle since by not appealing to the original position it is immune to any objections that are derived from decision-making procedures under conditions of (complete) uncertainty.[3]

The most important component of the argument and what ensures its distinctness is the introduction of the notion of bargaining and of determining an outcome or distribution that is based on the relevant parties reaching an agreement. The basic concept underlying the argument is not understood in terms of 'what mutually interested people choosing principles to advance their own interests from behind a veil of ignorance would agree in choosing,' (Barry 1989, 214) but rather in terms of what distribution of primary social goods rational agents would unanimously agree upon in the context of bargaining in a cooperative manner. The argument arrives at the difference principle by appealing to the Pareto principles which are subject to independent justifi-

[1] Thanks to _____[BLINDED]___

[2] Barry, for example, says that the Pareto argument is 'Rawls' fundamental argument for the difference principle' (1989, 213) and that it is 'the direct argument in chapter 2 of *A Theory of Justice*' (1989, 215).

[3] The most famous objection we have in mind here is John Harsanyi's derivation of utilitarianism from decision-making under conditions of complete uncertainty vis-à-vis the original position. See (Harsanyi 1982) for discussion.

cation and which are implemented in a particular bargaining model fit for the circumstances of justice. Thus, the notions employed in the argument are independent from the more well-known argument that appeals to the original position.[4]

Despite its uniqueness and *prima facie* strengths, the argument has been underdiscussed and has received little attention in the literature.[5] However, the argument has not gone unscathed. Two main criticisms of the Pareto argument are as follows. First, G. A. Cohen criticises the Pareto argument as being inconsistent in virtue of having one part that espouses equality as the initial *prima facie* just distribution of primary social goods (first stage of the argument) and another part that justifies a particular departure from this equality that is ultimately unequal (second stage). Cohen concludes that the egalitarian reservations that underpin the first stage cannot be upheld when the argument eventually tells us that an unequal outcome is nonetheless justified. Second, Patrick Shaw argues that '[t]he Pareto argument for inequality depends on taking it as self-evident that Pareto changes are changes for the better' (1999, 368). However, according to Shaw, it is not at all clear that Pareto changes are changes for the better and that the Pareto principles are self-evident. Thus, the grounds on which the Pareto argument depends are undermined and so the Pareto argument fails to get off the ground.

In this paper we revisit the Pareto argument, and examine and clarify it in detail (§1). In clarifying the argument we stress a subtlety that is overlooked in the little criticism it has received, namely, that the argument only justifies a particular unequal distribution which is the *most equal of the Pareto-optimal outcomes*. We argue that the justification for the Pareto principles derives from the principles of individual rationality and are not based on self-evidence, so, contra Shaw, the argument can get off the ground (§2). We further argue that the Pareto argument is not inconsistent because it is after all irrational to prefer the initial equal distribution in light of the most equal Pareto-optimal outcome, contra Cohen (§3). Finally we further develop the Pareto argument by explicating the kind of bargaining model required to arrive at the particular unequal distribution that the difference principle picks out. The model in question implements a principle, we call, 'full cooperation' and is not based on bargaining according to one's threat advantage but rather rational cooperation (§4).

§1: The Pareto Argument for Inequality

[4] David Lyons also detects this independent line of argument in Rawls, and thinks,

> The reason [why] is that one can construe Rawls' principles as a rational (self-interested) departure from an egalitarian norm, … This interpretation is encouraged by the fact that one can find in Rawls the suggestion of an argument for egalitarianism plus another argument, on rational (self-interested) grounds, to depart from that norm (1989, 152).

[5] The literature on the Pareto argument is contained within the following handful of articles/chapters: (Barry 1989, ch. 6; Cohen 1995; 2008, ch. 4; Shaw 1999).

The Pareto argument states that a certain distributional inequality of primary social goods is required by justice. This particular state of affairs is the just state of affairs. In Cohen's discussion of Rawls and Barry, he presents the Pareto argument as consisting in two stages. The first stage says that equality of opportunity or income is the initial *prima facie* just outcome, and the second stage says that a certain departure from the initial equality is justified so long as *everyone* benefits from the shift from the initial equality to this *particular* inequality.

The logical structure of this is not entirely clear, but we understand Barry and Cohen as saying the Pareto argument has two separate arguments as parts which when conjoined result in the conclusion that a certain unequal state of affairs is justified and that it is rational for agents in the initial state of affairs to prefer this unequal outcome over their current situation and will subsequently all agree on this Pareto-optimal outcome when rationally interacting in the bargaining problem of justice.

The first stage of the Pareto argument, according to Rawls and Barry, is grounded in our intuitive notion of justice. We think that being free and equal persons and being at the outset of a society requires that there be an equal distribution of social primary goods.[6] This is so because the differences in talent, interests, and genetic endowments are 'morally arbitrary' and thus should not be given weight in the initial distribution of goods.[7] However, given that our initial distribution is equal, there are certain other distributions that, while unequal, are to the *benefit of everyone*. One of these unequal but Pareto-optimal outcomes is to be preferred over the initial distribution if everyone is better off in this unequal Pareto-optimal state of affairs.

The second stage of the Pareto argument is grounded in rationality or rational self-interest. If all parties gain from shifting from one state of affairs to another, then each party must prefer this shift on pain of violating the most basic principle of rationality. It is irrational to prefer the initial equal distribution in light of *this* Pareto-superior state of affairs. Thus, it is rational to prefer this unequal Pareto-superior state of affairs.

Note well that the conclusion is not that so long as everyone gains we ought to prefer *any* Pareto-superior distribution to the initial equality. For this would have the untoward consequence of embracing Pareto-superior distributions that contain tremendous gaps of inequality as just. For instance, the Pareto argument does not licence a shift from the initial equal distribution (stage 1) to a Pareto-superior distribution such that group A and B are both better-off than in the initial distribution but that B has eight times the amount of social primary goods than A. The Pareto argument, to be clear, concludes that we ought to prefer 'the most egalitarian of all the Pareto-optimal arrangements satisfying the requirement that everyone should gain from inequality'

[6] See (Rawls 1971, 73-4) and also (Rawls 1993, 281).
[7] More specifically, Rawls and Barry argue for 'democratic equality,' see (Rawls 1971, §§12-13) and (Barry 1989, §27). This first component of the argument is not controversial and so a lengthy discussion of it is inappropriate.

(Barry 1989, 227).[8] The argument identifies a particular unequal state of affairs or distribution, and for Rawls, this result is exactly what the difference principle picks out. Furthermore, this is why Barry thinks that the Pareto argument is another way to argue for Rawls' difference principle.

The Pareto argument can now be stated as follows:

[Stage 1]

(P1) If talents, endowments, interests, etc, at the outset are 'morally arbitrary,' then the initial distribution should be equal.

(P2) Talents, endowments, interests, etc, at the outset are 'morally arbitrary.

(C) Thus, the initial distribution should be equal.

[Stage 2]

(P1) There is an unequal Pareto-optimal distribution that i) is strongly Pareto-superior to the initial equal distribution, and ii) is the most equal of the Pareto-optimal distributions.

---

[8] A major source of confusion, we believe, stems from conflating the Pareto argument with similar arguments that also implement Pareto principles. Below are two similar arguments that conclude that inequality is justified:

(P1) Either the distribution of primary social goods is equal or unequal.

(P2) The best feasible equal distribution is Pareto-inferior to some unequal distribution.

(C) Thus, an unequal distribution of primary social goods is always to be preferred (Cohen 1995, 163).

Obviously, this argument for inequality is not the Pareto argument because (P1) would be rejected by the first stage of the Pareto argument. The Pareto argument does not regard the relationship between equality and inequality as open-ended in this way. The Pareto argument insists on the initial distribution of goods being equal such that in at least one instance (but perhaps many more) an equal distribution is justified. But, (P1) is compatible with an initial distribution being unequal.

The second of the two arguments takes the first stage of the Pareto argument into account:

(P1*) Equal distribution of primary social goods is the *prima facie* just baseline.

(P2) The best feasible equal distribution is Pareto-inferior to some unequal distribution.

(C) Thus, an unequal distribution of primary social goods ought to be preferred.

Now, whilst this second argument is compatible with the first stage of the Pareto argument, it is non-specific with the *kind* of Pareto-optimal distribution that is to be preferred. This second argument is compatible with *any* Pareto-improvement from the initial equal distribution, even distributions that exemplify a tremendously large gap of equality between the better-off and worst-off. By contrast, the Pareto argument does not permit any Pareto-improvement from the initial equal distribution but rather permits one unique outcome, namely, the most equal of the Pareto-optimal distributions. Cf. (Rawls 1971, 69).

(P2) If there is an unequal Pareto-optimal distribution that i) is strongly Pareto-superior to the initial equal distribution, and ii) is the most equal of the Pareto-optimal distributions, then this unequal distribution ought to be preferred over the initial equal distribution.

(C) Thus, this unequal Pareto-optimal distribution ought to be preferred over the initial equal distribution.

There is one objection of Barry's reconstruction of Rawls that is worth dispelling here. One of Shaw's criticisms is to argue that it is unlikely that Rawls used the Pareto argument at all. Shaw presents the Pareto argument by quoting the following passage from *A Theory of Justice*:

> Imagine, then, a hypothetical initial arrangement in which all the social primary goods are equally distributed: everyone has similar rights and duties, and income and wealth are evenly shared. This state of affairs provides a benchmark for judging improvements. If certain inequalities of wealth and organizational powers would make everyone better off than in this hypothetical starting situation, then they accord with the general conception (Rawls 1971, 62).[9]

Shaw thinks this is the Pareto argument 'in a nutshell' and that the argument concludes that 'a position of inequality is justifiable so long as everyone benefits from it' (Shaw 1999, 354). From this, Shaw argues that 'it is doubtful whether Rawls does in fact employ the Pareto argument for inequality' (Shaw 1999, 353). Rawls' adoption of the difference principle rules out Pareto-improvements that exemplify a tremendous gap between the worse-off and better-off, and so Rawls would not accept every Pareto-improvement that is Pareto-superior to the initial equal distribution. However, according to Shaw, this would-be rejection by Rawls shows that Rawls does not employ the Pareto argument as friends of the Pareto argument accept that every Pareto-improvement is an improvement that is a 'change for the better'.

But this claim is misguided by the fact that Shaw fails to identify the Pareto argument that Barry reconstructed from Rawls. At the end of §11 from which the above passage was taken Rawls goes on to say that,

> [i]t is obvious, however, that there are indefinitely many ways in which all may be advantaged when the initial arrangement of equality is taken as a benchmark. How then are we to choose among these possibilities? The principles must be specified so that they yield a *determinate* conclusion (Rawls 1971, 65, our italics).

[9] NB: Rawls is referring to the *general* conception of justice which states that *all* 'social values—liberty and opportunity, income and wealth, and the bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage' (Rawls 1971, 62). This is to be contrasted with the *special* conception of justice which has a lexical (serial) ordering such that certain basic liberties cannot be exchanged for economic and social gains.

Rawls then goes on to consider what that determinate conclusion is in the next two sections. Rawls sets up the issue as determining which Pareto-optimal outcome is the just outcome as follows,

> 'The problem is to choose between them, to find a conception of justice that singles out one of these efficient distributions as also just. If we succeed in this, we shall have gone beyond mere efficiency yet in a way compatible with it' (Rawls 1971, 71).

The determinate conclusion, according to Barry's interpretation, is the most equal of the Pareto-optimal outcomes and it is only this outcome and the particular departure from the initial equality to this outcome that is justified. This is the determinate conclusion of the Pareto argument. Shaw's claim that Rawls did not employ the Pareto argument is simply incorrect insofar as Shaw has misidentified the structure of the Pareto argument.[10]

§2: The Justification of the Pareto principles

Shaw's overall objection against the Pareto argument says that the Pareto argument fails to get off the ground because it depends on taking 'as self-evident that Pareto changes are changes for the better,' but that, according to Shaw's objection, the claim that Pareto changes are self-evident is unjustified. So, the grounds on which the Pareto argument depends is undermined. However, as we will argue in this section, the premise that the Pareto argument depends on the intuitive proposal that the Pareto conditions are self-evident is false. The Pareto argument need not appeal to self-evidence or intuition in determining what changes are Pareto changes nor in grounding the Pareto principles proper. The appeal to self-evidence fails and rightly so, but it does not follow that the Pareto principles are left unjustified and that the Pareto argument fails to get off the ground. In fact the Pareto conditions are grounded in rational self-interest and so ought to be interpreted as conditions of individual rationality, or in other words, the Pareto conditions ought to be included among the basic set of principles of individual rationality which are concerned with individual preference and maximisation of expected utility. We will argue for this proposal in this section.

[10] One could object here and say that this is merely a terminological issue between us and Shaw. He has in mind one kind of argument that uses the weak-Pareto-principle (he calls it WPP for short) to justify any or several unequal outcomes, and we are thinking of a more fine-grained argument that attempts to justify a specific (the most equal of the Pareto-optimal) unequal outcome. This objection, we say, is incorrect. This is not a terminological issue insofar as Shaw takes himself to be responding to Barry's reconstruction of Rawls' argument in chapter 2 of *A Theory of Justice* and Cohen's critique of Barry. If anything Shaw is simply dialectically confused and his discussion of the Pareto argument severely suffers from this confusion. If Shaw thinks he is discussing the same argument as Cohen, then Cohen is also attacking a straw-man, and so much the worse for both of them. We think that Cohen can be interpreted as correctly representing Barry's reconstruction of Rawls although Cohen does not explicitly mention that a unique Pareto-optimal outcome is the determinate conclusion of the Pareto argument.

The Pareto principles,[11] (sometimes called definitions or conditions) are the core components of the Pareto argument which are used to derive the conclusion that a particular unequal distribution or outcome is justified.[12] The core principles (from note 11) viz., *(weak)Pareto-principle* and *(strong)Pareto-principle*, in plain English are as follows. If a change benefits everyone, then it is a change for the better, and if a change benefits some but harms no one, then it is a change for the better. Counterfactually, were a shift to a feasible Pareto-improvement in which everyone or some gain (but none are worst-off) not made, then it would be a bad outcome or a change for the worse.[13]

But, how exactly are the Pareto principles justified? One way to ground the Pareto principles is to appeal to their obviousness or the fact that they seem to be self-evident. For example, when considering a shift from one overall state of affairs to another such that one person benefits and no one else is made worse-off it seems intuitively compelling to say that it is a shift worth accepting or at least a shift that does not on the face of it seem objectionable. That the Pareto conditions seem intuitively compelling is made more salient when considering state of affairs in which *everyone* gains from shifting from one state of affairs to another. Why would anyone object to such a shift insofar as it is *ceteris paribus* a 'change for the better'? Such considerations can be thought of as instances of the more general claim that if someone can benefit from something without harming anyone else, then we ought to let such a state of affairs be realised. The self-evidence of which stares us in the face. Call this the *intuitive proposal*.

---

[11] Curiously named after the Italian economist Vilfredo Pareto (1848-1923).

[12] More formally, *(strong)Pareto-superiority*: if everyone is better off in S1 rather than S2, then S1 is strongly-Pareto-superior to S2. *(weak)Pareto-superiority*: if at least one person is better off and no one is worse off in S1 rather than S2, then S1 is weakly-Pareto-superior to S2. The contrary of Pareto-superiority being: *Pareto-inferiority*: if S1 is Pareto-superior to S2, then S2 is Pareto-inferior to S1. Further, if some state of affairs is Pareto-superior to S1, then S1 is Pareto-inferior (*tout court*). Next, we can characterise a specific Pareto-superior state of affairs that yields an optimal distribution also in two corresponding ways: *(strong)Pareto-optimality*: if no state of affairs is weakly-Pareto-superior to S1, then S1 is strongly-Pareto-optimal. *(weak)Pareto-optimality*: if no state of affairs is strongly-Pareto-superior to S1, then S1 is weakly-Pareto-optimal. In addition, the conditions under which certain states of affairs can be compared are as follows: *Pareto-comparability & -incomparability*: if either S1 or S2 are Pareto-superior to the other, then S1 and S2 are Pareto-comparable. However, if neither S1 nor S2 are Pareto-superior to the other, then S1 and S2 are Pareto-incomparable. If any two states of affairs are Pareto-comparable, then the *relation* between the two states of affairs can be characterised as: *(strong)Pareto-improvement*: if C1 benefits everyone, then C1 is a strong-Pareto-improvement. *(weak)Pareto-improvement*: if C1 benefits at least one person and no one is made worse-off, then C1 is a weak-Pareto-improvement. *(strong)Pareto-principle*: if C1 is a feasible weak-Pareto-improvement, then the strong-Pareto-principle mandates C1. *(weak)Pareto-principle*: if C1 is a feasible strong-Pareto-improvement, then the weak-Pareto-principle mandates C1.

[13] We are concerned with states of affairs in which everyone gains and so are concerned with (weak)Pareto-optimality. However, we think it is a shame that the ways of stating Pareto-optimality are usually characterised as weak and strong. It is misleading to call weak-Pareto-optimality weak since it requires the stringent condition that everyone gains. Some, we think, call it weak due to its obviousness, but as will be argued later we reject grounding the Pareto principles on intuition or self-evidence. A better label would be strict-Pareto-optimality, see (McClennen ms, Chapter 3: Rational Cooperation) and McClennen (forthcoming).

The intuitive proposal is put forth by several defenders of the Pareto principles, for example, economists, such as, Yew Kwang Ng and James Griffin. Griffin writes,

> there are standards for the good of many persons that are as obvious and inevitable as the standard for one person. There is this one for instance: *if some in a group become better off and none become worse off, the state of the group is better*. It is hard to see how anyone could resist such a principle (Griffin 1986, 147, original italics??).

The Pareto principles, according to this proposal, are justified by the fact that they are self-evident, intuitively compelling and downright obvious to the extent that to reject them requires a counter-intuitive claim not worth accepting without independent motivation and further argument. Kwang Ng goes so far as to say that to reject the Pareto conditions 'takes a rather peculiar ethic' (1980, 31). However, there are three objections against this kind of justification which we think provide at least *prima facie* reasons for rejecting the intuitive proposal.[14]

First, one can press hard on the claim to self-evidence by appealing to cases where it is morally questionable which harms we should take into account when determining whether a shift from one state of affairs to another is after all a Pareto shift. For instance, if we shift from one state of affairs to another in which the rich receive an increase in wealth and the worse-off remain at the same level of income, the Pareto conditions licence such an improvement. However, the very fact that there is an inequality of this nature in the state of affairs to which we shifted might reduce the well-being of the worse-off.[15] If there is such a negative effect in this shift, then friends of the Pareto conditions might claim that after all this is a shift worth accepting since the worse-off were not made even worse off. However, if we respond in this manner, we are taking a stand on what is to be counted as 'morally significant' when accounting for the well-being of everyone in the shift from one state of affairs to another (Shaw 1999, 361). Thus, determining whether a shift is a Pareto shift involves complicated moral issues that leave the justification of Pareto conditions far from being self-evident in the sense that it is unquestionably obvious. Put briefly, the fact that the Pareto principles assume a 'metric' by which to determine when some group or person is better off is not uncontroversial so as to secure the claim that the Pareto principles are simply self-evident or downright obvious.[16]

---

[14] We are largely in agreement with Shaw regarding the status of the intuitive proposal. As will be argued below however we do not think that friends of the Pareto conditions need to appeal to self-evidence to justify the Pareto conditions.

[15] George Smith argues that material inequality reduces the well-being of the worse-off so that relative poverty in addition to absolute poverty needs to be taken into account. See (Smith 1996), from (Shaw 1999, 360).

[16] One could response to the first objection by arguing that the claim to self-evidence is not undermined as such and that the first objection only shows that there are some cases where the Pareto conditions are overridden by other considerations. It would only be a matter of refining the application of the Pareto conditions by introducing a *ceteris paribus* clause to get another this objection. We think that this re-

Second, a problem can be posed against the Pareto conditions directly by showing that there are cases in which the Pareto principles are counter-intuitive or provide us with a counter-intuitive result. Shaw presents the following example that involves the supposed violation of one's rights:

> Unknown to Angela, her unscrupulous uncle has died and she stands to inherit enormous wealth. Knowing that Angela would immediately destroy the bequest because she disapproves of the life her uncle led, and knowing also that there is no risk that his actions will come to light, Angela's lawyer surreptitiously diverts the whole bequest to a cause from which everyone benefits (1999, 362).

Given that everyone benefits from the lawyer's actions, we have a Pareto improvement from one state of affairs to another. But it seems as though the lawyer violated the rights of Angela by disposing of her inheritance despite the fact that it benefited everyone. In response to this example the defender of the Pareto principles can bite the bullet and say that the shift that resulted from the lawyer's actions was after all a Pareto shift, one that is mandated by the Pareto principles. However, this response ends up being a contentious claim that very few will accept and opponents thereof will not accept. Thus, the Pareto principles do not seem to be self-evident or as obvious as some make them out to be.

The third objection is a general objection regarding claims to self-evidence concerning the justification of normative propositions. The general lesson being that the very appeal to self-evidence itself is epistemically insecure since people's intuitions clash on such matters and in many different places. Any purported justification based on self-evidence or intuition of normative principles which include the Pareto principles is ultimately doomed. In light of this, we advocate the justification of normative principles in instrumental terms that focuses on what people ought to *do* rather than what they ought to *believe*.[17] This way the Pareto principles are justified by appealing to rational self-interest and general principles of individual rationality. Such a justification we say is epistemically secure in that it is grounded in such means-ends reasoning governing what one *ought to do* given what one *ought to prefer*.

We say that the Pareto principles are grounded in rational self-interest and so the Pareto principles ought to be understood as principles of individual rationality. Hausman and McPherson put the proposal nicely,

> A Pareto optimum (also called a "Pareto efficient" allocation) is typically defined as a state of affairs in which it is impossible to make anyone better-off without making someone worse-off, but this purported definition is misleading.

---

sponse only weakens the strength of the Pareto conditions and raises new questions about the content of such *ceteris paribus* clauses which may also in the end be controversial. The point of this first objection would still stand. The Pareto conditions are not unquestionably obvious to award them the status of being self-evident nor can one *ground* them in such a claim. Thanks to ___[BLINDED]___.

[17] See (McClennen 2010; ms, Introduction) for discussion.

> It is more accurate to say that *R* is a "Pareto improvement" over *S* if nobody prefers *S* to *R* and somebody prefers *R* to *S*; then *R* is a Pareto optimum if and only if there are no Pareto improvements over *R* (Hausman & McPherson 2006, 65).

The point to take well is that it is misleading to characterise the Pareto conditions wholly in terms of global states of affairs that are related in various ways, instead it is clearer to characterise them in terms of the preferences of individuals. If the Pareto principles are understood as conditions of individual rationality, then we can directly state the conditions in terms of the agent's individual preferences. Call this the *individual proposal*.

In axiomatic models of bargaining used in economic theory,

> [Pareto-optimality] can be thought of requiring that the players collectively should behave in a rational way, since it specifies that the solution will select an outcome such that no other feasible outcome is preferred by all the players. The outcome must be a maximal element of the "social preference" defined by the intersection of all the individual preferences (Roth 1979, 11 [pdf]).[18]

Collective rationality seems to amount to the convergence of the individual preferences of every rational agent that is involved in the bargaining problem. John Nash himself claims that the assumption of Pareto-optimality 'expresses the idea that each individual wishes to maximize the utility to himself of the ultimate bargain' (Nash 1950, 159).[19]

The relevant Pareto conditions are listed below with our revised interpretation in light of the individual proposal (see also note x):

> *(strong)Pareto-improvement**: S1 is a strong-Pareto-improvement over S2 if 1) S1 is strongly-Pareto-superior to S2, and 2) everyone prefers S1 to S2.
>
> *(weak)Pareto-optimality**: S1 is weakly-Pareto-optimal iff there are no strong-Pareto-improvements over S1.
>
> *(weak)Pareto-principle**: if C1 is weakly-Pareto-optimal* and a strong-Pareto-improvement*, then the weak-Pareto-principle* mandates C1.

If everyone benefits from a shift to S1 which is strongly-Pareto-superior to S2, then everyone ought to prefer S1 to S2. For agent A to not prefer S1 to S2 would result in a

[18] Roth says elsewhere, '… since it specifies that the solution will select an outcome having the property that no other feasible outcome is preferred by all of the players' (1980, 308).

[19] See (Diskin & Felsenthal 2007) for a more recent axiomatic model. They modify Roth's axiom of Individual Rationality so as to include the condition that 'a player should not be considered 'individually rational' if he accepts an agreement that provides him with a utility lower than the minimal utility he can derive in case the parties reach a Pareto optimal agreement' (Diskin & Felsenthal 2007, 26). This in effect reincorporates the axiom of Pareto-optimality into the axiom of Individual Rationality.

violation of maximisation of expected utility. If everyone is better off in S1 rather than S2, then everyone ought to prefer S1 to S2. Further, if S1 is strongly-Pareto-superior to S2, then everyone ought to prefer S1 to S2. Ultimately we argue that in the case of justice there is one unique Pareto-optimal outcome that is to be preferred by all in the form of a unanimous agreement determined by a bargaining process between all rational agents cooperatively engaging in the distribution of primary social goods. Before we can fully develop this idea we need to defend the Pareto argument from Cohen's critique.

§3: Cohen's Critique of the Pareto Argument

Cohen argues that there is an inconsistency between the first and second stage of the Pareto argument. He claims that if one accepts 'equality of opportunity' and regards it as *prima facie* just, then one cannot accept 'the particular type of Pareto-improvement favoured by Rawls' (Cohen 1995, 162).[20] There is an inconsistency here because it is always possible for there to be a 'Pareto-optimal equal distribution that is also Pareto-superior to the initial equality' (1995, 162). If one has already accepted the initial equality, Cohen argues, one must prefer this Pareto-superior equal distribution over the Pareto-superior distribution of inequality suggested by Rawls' difference principle. To do otherwise would undermine the reasons for accepting the initial equality of the first stage of the argument. In other words, accepting the second stage of the Pareto argument is inconsistent with accepting the first.

Cohen has a more particular objection to the Pareto argument that rejects, what we will call, following Cohen, the *Rawlsian irrationality thesis*, which says that it is irrational prefer the initial equality in light of a particular unequal Pareto-optimal outcome.[21] Again, following Cohen, we call the initial equal distribution of social primary goods 'D1' and the particular unequal Pareto-superior distribution under consideration 'D2'. We will further stipulate that there are two groups of people in these and further states of affairs below known as the untalented 'A' and the talented 'B'. In D1, A and B have an equal number of primary social goods, however, in D2, the distribution of primary social goods is such that B (let's say) has a greater quantity than A. Cohen invites us to consider the following logically possible state of affairs 'D3'. D3 is strongly-Pareto-superior to D1 but A and B have an equal distribution of primary social goods. Cohen writes, '… D3 preserves equality, and the untalented are better off in D3 than they are in D2, while the talented are less well off in D3 than in D2, and

[20] This article is incorporated into Chapter 4 of Cohen's *Rescuing Justice and Equality*. We refer throughout to the original (1995) article.

[21] Rawls summarises much later that, '… an equal division of all primary goods is irrational in view of the possibility of bettering everyone's circumstances by accepting certain inequalities' (Rawls 1971, 546).

both are better off in D3 than they are in D1' (1995, 171). We represent this as follows:

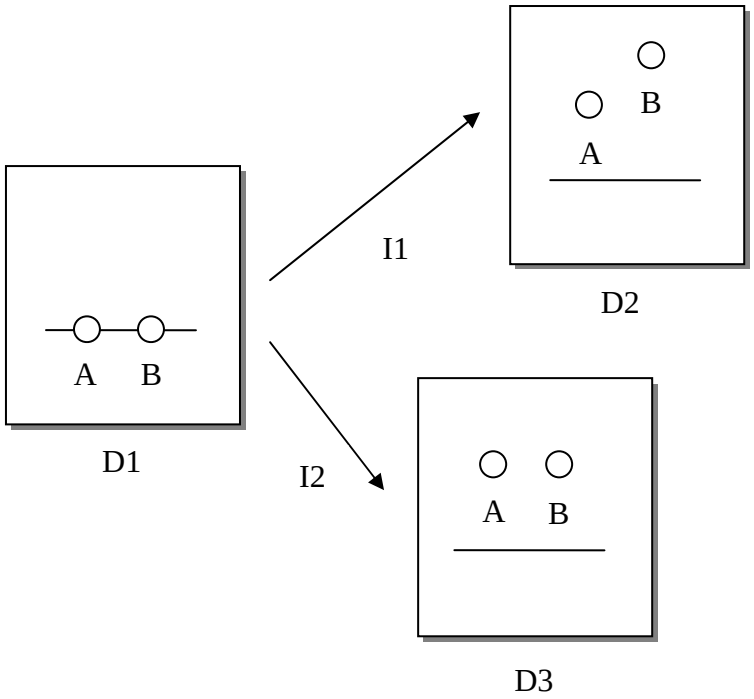

Figure 1

Cohen argues that if D3 is feasible (i.e., I2 is available) and B is willing to enjoy a lower wage rate that what they would enjoy in D2, then 'Rawls' claim about the irrationality of insisting on equality in the face of the possibility of a Pareto-superior inequality would lose its force, since a Pareto-improving *equality-preserving* move, in which no one is as badly off as some are in D2, would now also be available' (1995, 172, original italics). This more particular objection lends weight to the general objection that the Pareto argument is inconsistent as it attempts to show that there is no reason to say that it is irrational to prefer a shift from the initial equal distribution to a Pareto-superior distribution that is also equal. Call the more particular objection the *irrationality objection,* and the more general objection the *inconsistency objection*.[22] In this section, we argue that both objections do not succeed.

The inconsistency objection originates from a more general worry about how Rawls can say in the same breadth that an initial equality is just but that an inequality is justified even if everyone benefits. Cohen is not the first to pick up on this tension.[23]

[22] There is the question whether the more general objection, according to Cohen, depends on the more particular objection so that to argue against the more general we need only attack the more particular. However, as we will see, the more general objection has been raised by earlier commentators of Rawls which leads us to regard it as independent and worthy of separate discussion from the particular objection. We hope that our discussion of the more general objection then singles out and isolates the more particular objection which we then argue is also mistaken.

[23] See (Grey 1973) and (Narveson 1976, 1978).

Although Cohen fails to mention this, Barry does anticipate his inconsistency objection explicitly and discusses it at length.[24] The general objection, according to Barry, is that if we accept the initial equality as *prima facie* just, then any incentives become redundant as everyone would simply agree that it is just to enjoy the same wage rate. B agrees to enjoy a lower wage rate in D3 than in D2, because it is just to enjoy the same wage rate as A. So the purported shift from the initial equality to any Pareto-superior inequality is exacerbated since the reasons for regarding D1 as just are the same reasons for regarding D3 as just. Thus, we can only shift from the initial equal distribution to an equal Pareto-superior distribution. Ultimately justice demands of us an equal distribution.

The problem for Rawls then is how people who are committed to justice can have self-interested motivations for personal gain. If we claim that people need material incentives to produce according to their optimal capacity, then they are less concerned with the considerations and circumstances of justice and more fixated on personal gain. Barry first considers a response to this problem which demotes the difference principle to the level of being a second-rate principle or a principle of relative justice.[25] However this is not something Rawls would accept. He thinks that the difference principle is a principle of ideal justice and that ultimately a particular form of inequality is justified and ought to be preferred by all.

In order to explain the move from initial equality to Pareto-superior inequality and retain Rawls' claim that the difference principle is ideally just Barry argues that "freedom of occupational choice" is best understood as a "basic liberty" that is then subsumed under the first principle of justice and subsequently given priority over the difference principle. If Rawls is entitled to this move in light of his egalitarian premises, which Barry thinks he is, then the alleged inconsistency between the first stage and second stage of the Pareto argument is explained away. Barry concludes,

> 'I can therefore see no problem of internal coherence for Rawls' theory if we understand the application of the difference principle to be limited by the constraint of respecting freedom of occupational choice' (1989, 400).[26]

This qualification would also rule out Cohen's claim, which is at work in the irrationality objection, that B is willing to enjoy the same wage rates as A in D3 when B is

[24] See (Barry 1989, 396); cf. (Barry 1989, 234).

[25] If we accepted such, Barry writes, [i]t would then enable us to explain quite easily how we get from the premise that all inequalities based on differential productive ability are morally arbitrary to a conclusion in which there are inequalities corresponding to productive contributions. We would simply say that these inequalities are *not ideally just*, but that, once we concede the need for incentives, inequalities permitted by the difference principle are the only defensible ones (Barry 1989, 398, our italics).

[26] Rawls himself remarks, 'Thus justice is defined so that it is consistent with efficiency, at least when the two principles are perfectly fulfilled. Of course, if the basic structure is unjust, these principles will authorize changes that may lower the expectations of some of those better off; and therefore the democratic conception is not consistent with the principle of efficiency if this principle is taken to mean that only changes which improve everyone's prospects are allowed' (Rawls 1971, 79).

working more hours than A. Since considerations of justice do not demand that B accept or be willing to accept the same (and hence lower) wage rate as A, such a constraint can only be established by 'some combination of coercion and moral suasion' (Barry 1989, 400). But Rawls and the proponent of the Pareto argument need not accept this kind of coercion or moral suasion. More importantly, we ought to remind ourselves that the conclusion of the Pareto argument coincides with the outcome that the difference principle picks out. So when parties are bargaining in a particular cooperative setting, namely, the bargaining problem of justice, they have already secured certain sets of rights prudentially.[27]

Recall that the irrationality objection says that the irrationality of preferring an equal distribution to an unequal distribution is undermined when one considers the possibility of D3—an equal distribution that is Pareto-superior to D1. In short, Cohen is arguing that in light of I2 (i.e., the shift from D1 to D3) it is not irrational to prefer an equal distribution over an unequal distribution. However, the general claim that it is not irrational to prefer any equal distribution over an unequal one is false. By Cohen's lights, B is worse off in D3 than in D2 despite the fact that Cohen makes it the case that B is willing to enjoy the same wage rate as A. If D3 is feasible, then there is a Pareto-superior unequal distribution, call it D4, in which B is better off and A is not made worse off. This in effect renders D3 as a sub-optimal state of affairs even though it is Pareto-superior to D1. The shift from D3 to D4 tells us that it is irrational to prefer an equal distribution over an unequal one. B would prefer D4 to D3 and A would be indifferent between the two and thus weakly prefers D4 to D3.[28] Thus, Cohen has failed to undermine the Rawlsian irrationality thesis.

§4: Conclusion

Although we have shown that Shaw is discussing a straw-man and not the Pareto argument *per se*, he does have a well-taken criticism of Cohen that is worth mentioning. To be clear, Rawlsian exegesis is our least concern, and we will suppose for the purposes of explicating Shaw's objection that the conclusion endorsed by the Pareto argument is that *any* unequal Pareto-superior distribution is justified. We will then discuss Shaw's second response to Cohen that helps illuminate his misinterpretation of the Pareto argument.

[27] See (McClennen ms, chapter on Prudence and Rights).

[28] Cohen mentions the response that if we have D3 then why don't we start with D3. He asks, 'Why was the original equality not pitched at D3, instead of at D1, when D3 is Pareto-superior to D1?' (1995, 174). But then says nothing else about this response other than saying that '[h]ad we begun with D3, D2 would have been seen for what it is'. This is a red herring however. If we had started with D3, then the unequal Pareto-superior distribution in which a strong Pareto-improvement is available would not be D2 but rather some other Pareto-superior distribution.

Shaw attempts to show that Cohen's purported solution to the tension between the two stages of the Pareto argument 'ultimately sidesteps the problem'. According to Shaw, Cohen attributes to Rawls two 'value judgements' which conflict: 'that departures from equality are unacceptable, and that changes from which everyone benefits are for the better' (Shaw 1999, 358). Cohen then argues that the equality-preserving Pareto-superior state of affairs D3 does not offend either of these 'value judgements'. But, Shaw asks, how is Cohen to evaluate I1 (i.e., the shift from D1 to D2) when D3 is feasible, or better, when I2 (i.e., the shift from D1 to D3) is available? If we follow the weak Pareto principle, I1 is to be regarded as a 'change for the better,' however by establishing that justice ultimately requires an equal distribution, we are forced to say that I1 is a 'change for the worse'. But this contradicts the spirit of the weak Pareto principle which claims that any unequal Pareto-superior distribution is a 'change for the better'. Thus, the fact that Cohen points out that D3 is feasible does not eliminate the alleged tension between the two stages of the argument.

Insofar as this is an objection against Cohen's positive claim we adopt the point as our own. Cohen does not eliminate the tension between the two stages of the argument by simply pointing out the feasibility of D3 or the availability of I2. However, as we mentioned earlier, Shaw's exposition of the argument is strictly speaking mistaken. Whilst it is true that the weak Pareto principle says that 'a change for the better' results from *any* unequal Pareto-superior distribution to the initial equality, Rawls and Barry do not *implement* the Pareto principles in this way. They understand, of course, that there are many unequal distributions that are morally repugnant and ought not to be considered when addressing the circumstances of justice.

To explain how Shaw has misunderstood the Pareto argument we will discuss the other worry he has with Cohen's solution which he thinks involves 'a strange consequence'. For Cohen, if D3 is possible, then D2 is unacceptable. However, D1 is also unacceptable as well. D1 is unacceptable just because it is Pareto-inferior to D2 and D3. Shaw concludes,

> 'And thus Cohen's solution would commit Rawls to holding that *every* position is unacceptable except for the most equal feasible Pareto-optimal position' (Shaw 1999, 358, original italics).[29]

But this is exactly the conclusion that the Pareto argument demands. Recall that the current dispute dialectically speaking right now is about *the Pareto argument* and not Rawlsian exegesis; and as Barry explicitly argues this is the determinate conclusion. For example,

[29] One of Shaw's immediate reactions to this so called 'strange consequence' is that it is 'so high a standard of acceptability' (Shaw 1999, 358). But why, under circumstances of justice, should one think that a high standard is unwarranted? If anything we ought to retain such a high standard given what is at stake.

> 'It may be recalled that the difference principle was to satisfy the criterion that "everyone must gain" from an inequality in some sense *beyond* that of creating a Pareto improvement on equality' (Barry 1989, 233, our italics).

And again,

> 'The difference principle picks out the most egalitarian of all the Pareto-optimal arrangements satisfying the requirement that everyone should gain from inequality' (Barry 1989, 227).[30]

Shaw's 'strange consequence' can be interpreted as an objection against the Pareto argument even though he has failed to properly identify it. Shaw argues that it is a strange consequence because '… even if better situations than D1 are feasible, that would not normally be felt to rule out D1's being an acceptable, or a good, situation' (Shaw 1999, 358). But this is to miss the whole point of the Pareto argument and the fact that it proceeds in *stages*. Rawls also says that '[j]ustice is prior to efficiency and requires some changes that are not efficient in this sense' (Rawls 1971, 79-80). We first secure an initial equal distribution as *prima facie* just, and secure this on independent grounds stemming from the fact that endowments, talents, etc are morally arbitrary. Once we have secured this, then parties bargain to determine a Pareto-superior distribution to the initial distribution they have already secured and can fall back on if one of the agents does not prefer a proposed distribution. This bargaining process, which is the essence of the Pareto argument, proceeds *across time*. The initial equal distribution, whilst sub-optimal if other states of affairs are Pareto-optimal, is the 'point of no agreement,' and Individual rationality as an axiom in standard and non-standard bargaining models guarantees that agents can terminate any agreement and walk away with what they started with, and in the case of justice they can walk away with the initial equal distribution. However, as the argument goes, it is irrational to do so as the inequality that can be realised if everyone cooperates ensures that everyone is better off.

Moreover, it is misleading to say that every *position* is unacceptable 'except for the most equal feasible Pareto-optimal position' as many states of affairs will be Pareto-incomparable and thus such a judgement is not possible. It is more accurate to say that every *improvement* is unacceptable except for the improvement that shifts everyone to the most equal feasible Pareto-optimal distribution. The most equal outcome respects the justification of the first stage because it is concerned with the gap of inequality between the worst-off and better-off, and is concerned with how differing groups receive their gains. The most equal outcome is just if the gains had by the better-off do not come at the expense of the worst-off. The outcome is a cooperative one, and the bargaining between, say, A and B is one of cooperation, where all parties gain but not at

[30] Just to thrash this point, Barry announces his aims at the outset of chapter 6: 'I shall try to show that the difference principle does, as Rawls maintains, pick out a unique Pareto-optimal point that is a Pareto improvement on an equal distribution' (Barry 1989, 214).

the expense of anyone else. We reiterate that the Pareto argument focuses on cooperative outcomes where the distribution of social primary goods comes about when the parties or groups involved interact cooperatively. In these cooperative outcomes, the gains to the differing parties or groups are not realised at the cost of any other party or group. To be clear, 'everyone gains' really means 'everyone gains and no one loses'. This follows from the outcome being a cooperative one.

Bibliography

Barry, B., (1989) *Theories of Justice: A Treatise on Social Justice vol. 1* (London: Harvester Wheatsheaf).

Cohen, G. A., (1995) 'The Pareto Argument for Inequality,' *Social Philosophy and Policy* 12(1), 160-85.

Cohen, G. A., (2008) *Rescuing Justice and Equality* (Harvard: Harvard University Press).

Diskin, A., & Felsenthal, D., (2007) 'Individual Rationality and Bargaining,' *Public Choice* 133(1/2), 25-29.

Grey, T., C., (1973) 'The First Virtue,' *Stanford Law Review* 25, 286-327.

Griffin, J., (1986) *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Oxford University Press).

Harsanyi, J., (1982) 'Morality and the Theory of Rational Behavior,' in A. Sen & B. Williams (eds.) *Utilitarianism and Beyond* (Cambridge: Cambridge University Press), 39-62.

Hausman, D., & McPherson, M., (2006) *Economic Analysis, Moral Philosophy, and Public Policy* (2nd ed.)(Cambridge: Cambridge University Press).

Lyons, D., (1989) 'Nature and Soundess of the Contract and Coherence Arguments,' in N. Daniels (ed.) *Reading Rawls: critical studies on Rawls' A Theory of Justice* (Stanford, CA: Stanford University Press), 141-67.

McClennen, E. F., (2010) 'Rational Choice and Moral Theory,' *Ethical Theory and Moral Practice* 13(5), 521-40.

McClennen, E. F., (ms) *Rational Society* (book: manuscript).

Narveson, J. F., (1976) 'A Puzzle about Economic Justice in Rawls' Theory,' *Social Theory and Practice* 4, 1-27.

Narveson, J. F., (1978) 'Rawls on Equal Distribution of Wealth,' *Philosophia* 7, 281-92.

Nash, J., (1950) 'The Bargaining Problem,' *Econometrica* 18(2), 155-62.

Ng, Y., (1980) *Welfare Economics: the Introduction and Development of Basic Concepts* (New York: Wiley).

Rawls, J., (1971) *A Theory of Justice* (Cambridge, Mass.,: Belknap Press of Harvard University Press).

Rawls, J., (1993) *Political Liberalism* (New York: Columbia University Press).

Roth, A. E., (1979) *Axiomatic Models of Bargaining* (Dordrecht: Springer Verlag).

Roth, A. E., (1980) 'The Nash Solution as a Model of Rational Bargaining,' in A. V. Fiacco & K. O. Kortanek (eds.) *Extremal Methods and Systems Analysis* (Dordrecht: Springer Verlag), 306-11.

Shaw, P., (1999) 'The Pareto Argument and Inequality,' *The Philosophical Quarterly* 49(196), 353-68.

Smith, G., (1996) 'Income Inequality and Mortality: Why are They Related?,' *British Medical Journal* 312, 987-8.