ECINEQ

Society for the Study of Economic Inequality

Working Paper Series

Prioritarianism and Equality of Opportunity

Paolo Brunori

Francisco Ferreira

Vito Peragine

ECINEQ 2021 574

ECINEQ 2021 574
2021 February
www.ecineq.org

Prioritarianism and Equality of Opportunity

Paolo Brunori
University of Firenze & Bari
Francisco Ferreira
London School of Economics
Vito Peragine
University of Bari

Abstract

This paper asks whether prioritarianism – the view that social welfare orderings should give explicit priority to the worse-off – is consistent with the normative theory of equality of opportunity. We show that there are inherent tensions between some of the axioms underpinning prioritarianism and the principles underlying equality of opportunity; but also that these inconsistencies vanish under plausible adjustments to the domains of two key axioms, namely anonymity and the transfer principle. That is: reconciling prioritarianism and equality of opportunity is possible but allowing room for individual responsibility within prioritarianism requires compromises regarding the nature and scope of both impartiality and inequality aversion. The precise nature of the compromises depends on the specific variant of the theory of equality of opportunity that is adopted, and we define classes of social welfare functions and discuss relevant dominance conditions for six such variants. The conflicts and the paths to reconciliation are illustrated in an application to South Africa between 2008 and 2017, where results suggest broad empirical agreement among the different approaches.

Keyword: Prioritarianism; welfarism; equality of opportunity; stochastic dominance; robust welfare comparisons; South Africa

JEL Cassification: I31, D63

Prioritarianism and Equality of Opportunity

Paolo Brunori, Francisco H. G. Ferreira and Vito Peragine[1]

26 January 2021

Abstract: This paper asks whether prioritarianism – the view that social welfare orderings should give explicit priority to the worse-off – is consistent with the normative theory of equality of opportunity. We show that there are inherent tensions between some of the axioms underpinning prioritarianism and the principles underlying equality of opportunity; but also that these inconsistencies vanish under plausible adjustments to the domains of two key axioms, namely anonymity and the transfer principle. That is: reconciling prioritarianism and equality of opportunity is possible but allowing room for individual responsibility within prioritarianism requires compromises regarding the nature and scope of both impartiality and inequality aversion. The precise nature of the compromises depends on the specific variant of the theory of equality of opportunity that is adopted, and we define classes of social welfare functions and discuss relevant dominance conditions for six such variants. The conflicts and the paths to reconciliation are illustrated in an application to South Africa between 2008 and 2017, where results suggest broad empirical agreement among the different approaches.

Keywords: Prioritarianism; welfarism; equality of opportunity; stochastic dominance; robust welfare comparisons; South Africa.

[1] Brunori is at the University of Florence and University of Bari; Ferreira is at the London School of Economics and IZA; Peragine is at the University of Bari. This working paper is a preprint (SMUR) version of the chapter "Prioritarianism and Equality of Opportunity" part of the book "Prioritarianism in Practice" edited by Matthew Adler and Ole Norheim and published by Cambridge University Press. We are grateful to Matt Adler, Ole Norheim, and participants at the PiP Network workshops at Duke University and in Bergen for helpful discussions. We are solely responsible for any errors.

1. Introduction

What objective, if any, should society – or the State as its collective agent – seek to pursue for its own betterment? What social objective function – again, if any – should well-intentioned policymakers seek to maximize, as they choose among multiple possible outcomes?[2] For about two centuries, one single (and surprisingly simple) answer to this question has been supremely dominant across the social sciences, namely utilitarianism. Normally attributed to Jeremy Bentham (1789), the modern formulation of utilitarianism posits that society is made up of a collection of individuals, each of whom has a well-defined level of well-being $w_i(x)$, which depends on outcome *x*. Society's objective, then, is to choose the feasible outcome *x* where the sum of the individual levels of well-being over the population is greatest.

The influence of utilitarianism over generations of philosophers and social scientists cannot be overstated. Speaking of economists, Sen (2000) writes: "...in many respects, utilitarianism serves as the 'default program' in welfare-economic analysis: the theory that is implicitly summoned when no others are explicitly invoked." (p.63).

Yet in the late twentieth century, the unrivalled dominance of utilitarianism began to be questioned, often from very different perspectives. Libertarians argued that society would be better served by ensuring that certain fundamental individual rights were respected, rather than by seeking to maximize some notion of utility or well-being. Rawls (1971) suggested his famous two principles of justice, which included an Equal Opportunity Principle. Sen (1980, 1985) himself argued for a broader view of people's capabilities as the basal space for his concept of social justice. And so on.

Many of these perspectives have been extensively reviewed elsewhere (e.g. Roemer, 1993 and Sen, 2000) and we do not dwell further on them here. Our narrower focus is on two among these alternative approaches that were critical of utilitarianism in its pure form, namely (what is now known as) prioritarianism, and the equality of opportunity approach. Prioritarianism is the name given by philosophers to an approach that incorporates the idea that inequality in the distribution of well-being is costly in terms of social welfare. That is to say: merely summing across individual well-being levels ignores an important dimension of the social objective. If two outcomes, *x* and *y*, yield the same sum of well-being across society, but well-being in *y* is distributed more unequally than in *x*, then prioritarianism would rank *x* as preferable to *y*.

[2] The word "outcome" is used here to denote alternative worlds, or model representations.

As we will briefly discuss in the next section, when a few other desiderata are taken into account, this implies that instead of summing across individual levels of well-being, one ought to sum across strictly concave transformations of well-being – so that an extra unit of well-being earned by a less well-off person contributes more to the total than if that unit had accrued to a better-off person. Formally introduced to philosophers by Derek Parfit (2000), this view corresponds directly to the notion of strictly concave and additive social welfare functions (SWFs), which were familiar to economists since at least the late 1960s and early 1970s.[3]

The normative theory of equality of opportunity (E.Op. for short), on the other hand, is driven by the idea that not all differences in well-being are normatively equivalent. Proponents postulate that inequality in well-being can arise because of circumstances beyond the control of individuals (such as a person's race, gender, parents, or birthplace) or because of the exercise of individual responsibility and effort.[4] They argue that the first kind of inequality is ethically unacceptable and should be compensated, whereas the second kind is permissible and does not warrant compensation from society.[5] This is perhaps a more substantive departure from utilitarianism, in that it questions not only the aggregation procedure (across individual levels of well-being), but also the very "basal space" (Sen, 2000) upon which normative judgments should be made. The argument is that society should promote greater (and less unequal) amounts of *opportunity for welfare* among the population, rather than focusing on the distribution of welfare itself.

While both of these normative approaches to social justice constitute departures from pure-form utilitarianism, they are clearly different from each other. Prioritarianism still relies exclusively on the space of well-being to assess and rank different potential social outcomes and, in this sense, it is still 'welfarist'. But it does so while incorporating an aversion to inequality in well-being. Equality of opportunity also incorporates a form of inequality aversion but, critically, that aversion applies only to some forms of inequality and not to others. The space of well-being is no longer sufficient for assessing and ranking social

[3] See, for example, Atkinson (1970).

[4] Seminal contributions to this literature were made by Arneson (1989), Cohen (1989) and Dworkin (1981a,b) among philosophers, and by Fleurbaey (1994, 1995), Roemer (1993, 1998) and van de Gaer (1993) among economists.

[5] In this paper, we will mostly gloss over the longstanding debate about how "luck" should be treated in this framework. At the risk of simplifying excessively, one may think of "brute luck" as being a stochastic component of the set of circumstances, and "option luck" as having at least some responsibility component.

outcomes: E.Op. requires additional information on the *sources* of well-being, and the role of individual responsibility among those sources. In this sense, it is a 'non-welfarist' approach to social justice.

The question we ask in this paper is whether these two non-utilitarian approaches to social justice are – or can be made – mutually compatible. A priori, it may seem obvious that the answer is "no". Indeed, Adler (2018), who considered a very similar question before us, comments: "The reader might observe that Conflict is obvious. '*Of course* it's true that introducing a non-well-being element into the goodness ranking will be inconsistent with the focus on well-being that the well-being Pareto principles embody – or so the reader might think." (p.26) But he goes on to argue that "the inconsistency is not [in fact] obvious", and to describe conditions under which it might not hold.

We follow a different path from Adler's and arrive at somewhat different results, but they are similar in one key respect: there are clear inconsistencies between the two approaches – as one would expect. As we will see, the inconsistencies are driven primarily by two of the fundamental axioms of prioritarianism: anonymity (also known as symmetry) and the Pigou-Dalton transfer principle. Yet we also show, drawing on earlier results in the literature, that suitable restrictions on the domain of those two axioms (along with some strengthening of the separability axiom) can make modified versions of prioritarianism consistent with the principles of equality of opportunity. Because there are different versions of those principles in the E.Op. literature, the nature of the ethical compromise differs according to which approach one subscribes to, and the empirical requirements differ substantially among them.

In what follows, we examine six different axiomatic "definitions" of equality of opportunity – arising from two alternative versions of the principle that inequalities due to circumstances (e.g. race, sex, family background, etc.) should be compensated, and three different attitudes to how individual effort and responsibility can be rewarded.[6] For each of the six definitions, we present a set of axioms that jointly define a family of social welfare functions embodying certain prioritarian properties. These families admit versions of the Atkinson SWF (or of Kolm-Pollack SWFs). Rather than focusing on one or two specific SWFs, we follow an alternative approach: for each family of SWFs defined by the different sets of axioms, we present the conditions under which a pair of distributions will be ranked unanimously by all members of that SWF family. This approach uses stochastic dominance relationships between two distributions (or

[6] One of these six "cells" has been approached in two difference ways in the literature, so a seventh empirical comparison is discussed in the Appendix.

functionals[7] thereof) to provide robust welfare rankings: if those conditions hold, then all members of the family of welfare functions will rank the two distributions the same way. It goes back to Atkinson's (1970) results about the link between Lorenz dominance and welfare dominance, and various generalizations thereof (e.g. Shorrocks, 1983). The approach is standard in welfare economics, but we briefly summarize it below and provide references for the unfamiliar reader.

The remaining of this paper is organized as follows. The next section lays out a basic analytical framework for thinking about this question and provides a brief overview of prioritarianism and the theory of equality of opportunity, including the distinction between ex-ante and ex-post approaches to E.Op.. Section 3 describes the conflict between prioritarianism and ex-ante E.Op., as well as the conditions under which the two can be reconciled by suitable axiomatic compromises. It presents the relevant dominance relationships allowing for robust rankings in each case. Section 4 does the same for ex-post E.Op., and Section 5 provides an empirical illustration using five waves of a panel household survey for South Africa, likely the most unequal country in the world today. Section 6 concludes. While some basic mathematical notation and a few key results are included in the main text, formal statements are consigned to footnotes or to Appendices A.1 and A.2.

2. The basic framework[8]

Consider a discrete population of fixed and finite size *N,* with individuals indexed by l. Denote individual well-being in outcome x by $w_l(x)$, and the population distribution of well-being by the *N*-dimensional vector ***w(x)***. Suppose $w_l(x)$ is a function of multiple personal attributes ***a**$_l$**(x)**,* and that this *(p+1)*-dimensional vector ***a,*** which fully determines $w_l(x)$, can be unambiguously divided into two kinds of attributes: those that are given exogenously to the individual, in the sense that they are beyond her control, ***C**$_l$**(x)**,* and those over which she can exert at least a modicum of control, $e_l(x)$. Let ***C**$_l$**(x)*** be a *p*-dimensional vector of circumstances that characterize each individual and $e_l(x)$ be a one-dimensional scalar index for effort or responsibility. Furthermore, denote $\varphi(x)$ as the set of tools ("policies") available to the policymaker in state *x* with which to influence well-being – say, by taxing some people and making

[7] A functional is simply a "function of a function", that is a functional relationship whose arguments may include other functions.

[8] Parts of this section draw on Ferreira and Peragine (2016), who discuss the "canonical" model of equality of opportunity in greater detail. The reader unfamiliar with the E.Op. literature is directed to that survey for a lengthier exposition.

transfers to others. Note that policies are chosen for society as a whole. They may vary among outcomes but, in each outcome, they are common to all.

We can then write individual well-being as:

$$w_l = w(C_l, e_l, \varphi) \qquad (1)$$

where each argument of the function *w(.)* is itself a function of state *x*, omitted to simplify notation.[9] For expositional simplicity, let *e(x)* and all elements of ***C(x)*** be discrete variables.[10]

Given *C* and *e*, the population can be partitioned in two ways. It can be divided into mutually exclusive groups of people who share identical circumstance vectors, C_i. In the literature, each of these circumstance-homogeneous groups is called a *type* (indexed by *i* and denoted T_i). Alternatively, we can partition the population into mutually exclusive groups of people who exert the same level of effort, e_j. These groups are called *tranches* (indexed by *j* and denoted T^j). (1) can then be re-written as:

$$w_{ij} = w\left(C_i, e_j, \varphi\right) \qquad (1')$$

Since each individual l must belong to a single type and a single tranche, (1') merely rewrites the individual well-being function by indexing individuals by the type (*i*) and tranche (*j*) each one belongs to. If, without loss of generality, there are *n* types and *m* tranches then the population can be represented by a matrix $\left[W_{ij}\right]$, as in Figure 1. In this matrix, each row denotes the support of the well-being distribution of a type, whereas each column denotes the support of the well-being distribution of a tranche.[11] It will prove convenient to define effort *e(x)* so that well-being is always (weakly) increasing in effort within each type.[12]

[9] In this paper we write $w_{ij}(x) = w(C_i(x), e_j(x), \phi(x))$ to indicate that individual well-being depends on her circumstances and efforts, as well as on the set of policies in place in outcome x. We also abstract from uncertainty, so that each outcome corresponds to a single state of nature.

[10] In this discrete case, the well-being function is a mapping $w: \Omega \times \Theta \times \Phi \to \mathbb{R}$, where $\Omega \subseteq \mathbb{Z}^p$, $\Theta \subseteq \mathbb{Z}$ and $\Phi \subseteq \mathbb{R}$. Note that $w \in \mathbb{R}$, so differentiable functions of *w* can be defined. Discreteness of the circumstance and effort variables is not important for the analysis, but simplifies exposition.

[11] Distributions, like other functions, are mappings from a domain to a range. The support of a distribution is simply its domain: the set of all possible values that the variable being distributed can take. In this case, the distributions themselves are vectors where each w_{ij} value is entered p_{ij} times (see below).

[12] That is, the effort variable is <u>defined</u> so that $w_{ij} = w\left(C_i, e_j, \varphi\right) \leq w_{ij+1} = w\left(C_i, e_{j+1}, \varphi\right)$. An example illustrates: suppose Janet's well-being is increasing in the result of a school exam. If studying for twelve hours for the test and not sleeping leads to a worse exam result than studying six hours and sleeping for six hours, all else equal, then the second strategy involves the higher effort level.

In general, there may of course be more than one person of type T_i in tranche T^j (that is, in each cell in the $[W_{ij}]$ matrix), so that a full description of the population would require defining a population size matrix $[P_{ij}]$, which is also *n x m*, and whose elements p_{ij} give the number of people with circumstances C_i and effort e_j. Naturally, $\sum_{1=1}^{n} \sum_{j=1}^{m} p_{ij} = N$.

This simple "model" of a society contains all the elements we will need to investigate the inconsistencies between prioritarianism and E.Op.. In particular, the introduction of a vector of circumstances, ***C(x)***, and an effort index, *e(x),* will allow us to treat differences in well-being differently, depending on whether they are driven by each. The matrix representation of society in Figure 1 is only a little richer than the vector representation familiar from welfarism (where all the information required for a social evaluation is contained in the distribution of well-being itself). But it will allow us to compare the two approaches. We begin with prioritarianism.

Figure 1: circumstances, effort and well-being

	e_1	e_2	e_3	...	e_m
C_1	w_{11}	w_{12}	w_{13}	...	w_{1m}
C_2	w_{21}	w_{22}	w_{23}	...	w_{2m}
C_3	w_{31}	w_{32}	w_{33}	...	w_{3m}
...	...	...	...	...	...
C_n	w_{n1}	w_{n2}	w_{n3}	...	w_{nm}

Prioritarianism

A Benthamite utilitarian would view the sum of individual levels of well-being in society,

$$S^U = \sum_{l=1}^{N} w_l = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} w_{ij} \tag{2}$$

as the appropriate social objective function. For her, this should be the maximand – the object a benign social planner should choose policies φ to maximize. A prioritarian, on the other hand, rejects (2) on the ground that it is insensitive to the distribution of well-being. She would prefer to maximize:

$$S^P = \sum_{l=1}^{N} g(w_l) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} g(w_{ij}) \tag{3}$$

with $g'(w) > 0,\ g''(w) < 0$.[13] The strict concavity of the transformation function $g(w)$ ensures that a gain in well-being of Δw to a poorer person (in terms of well-being) makes a greater contribution to social welfare than an identical gain to a richer person. Indeed, Adler (2018) shows that if (and only if) a social planner holds to five key principles (plus a few more technical properties), then her social rankings across outcomes will be mirrored by some member of the class of social welfare functions described by (3). The five principles, or axioms, are as follows:[14,15]

I. *Anonymity (or symmetry):* If an outcome *y* is obtained from another outcome *x* merely by re-arranging well-being levels w_l among individuals (so that ***w(y)*** is a permutation of ***w(x)***), then one should be indifferent between *x* and *y*.

II. *Strong monotonicity (or strong Pareto):* If an outcome y is obtained from another outcome x by raising the well-being level of at least one person, and lowering no one's, then y is preferred to x.

III. *Pigou-Dalton Transfer Principle*: Suppose well-being levels are the same across outcomes *x* and *y* for all but two people. If y is obtained from x by means of a mean-preserving spread – that is a pure (non-leaky) transfer from the poorer to the richer person – then x is preferred to y. Equivalently, if y is obtained from x by means of a progressive rank-preserving pure (non-leaky) transfer – that is, from the richer to the poorer person, without switching their ranks – then *y* is preferred to *x*.

[13] For presentational purposes we assume twice-differentiability of the transformation functions $g(w)$, even though this property is not in fact required for any of the results reviewed in this paper (and thus not one of the axioms listed).

[14] These axioms are standard in welfare economics and we state them informally here. Formal statements can be found in many different sources, e.g. Peragine (2004) or Adler (2019).

[15] Our earlier assumption that the population size is fixed at N across all possible outcomes implies that we can dispense with the Population Replication Invariance axiom, which would otherwise be needed. If N were variable and that axiom were imposed, then (2) and (3) would involve a division by N, yielding the more frequent "per capita" interpretation.

IV. *Separability*: The contribution of any individual to social well-being is independent from the contributions of others.[16]

V. *Continuity*: If state x is preferred to y given the two N-dimensional vectors *w(x)* and *w(y)*, then there exists a N-dimensional vector $\varepsilon \neq 0$ such that $w(x) \pm \varepsilon$ is also preferred to y.

Although these axioms are well-known to readers familiar with welfare economics or political philosophy, it is worthwhile commenting briefly on each. Anonymity ensures the policy-maker's impartiality: All else equal, she is indifferent between a world where Anne has a well-being of 5 while Paul has 10, and another where Anne has 10 and Paul has 5. This axiom requires that *w* be the only argument of the transformation function, *g(w).* The Pareto principle (or strong monotonicity) rules out levelling down: social welfare improves if well-being rises for some while remaining unchanged for all others. It mandates the positive first derivative of the transformation function.

The Pigou-Dalton transfer principle is the locus of inequality aversion in the prioritarian social welfare formulation: it requires that a mean-preserving spread – a transfer from a poorer to a richer person – reduce social welfare. Equivalently it requires that a rank-preserving transfer from a richer to a poorer person increases social welfare: so long as the total amount of well-being is unchanged, a less unequal distribution is always preferred. This axiom mandates the negative second derivative – that is, the strict concavity – of the transformation function. Separability requires that the change in a person's well-being level affects social welfare only directly, and not through the well-being of others. A mathematical formulation says that the cross-partial derivative of S^P with respect to the individual wellbeing of two distinct individuals, w_l and w_k, is zero. This is by no means a trivial requirement, and it is not invulnerable to criticism. The reason it is ubiquitous in welfare economics is that it mandates the additive formulation of (3), which makes analysis more tractable. Continuity is in a sense a more technical requirement, but it plays the important role of preventing very small changes in a person's well-being from having a disproportionately large effect on social welfare. It mandates the continuity of the transformation function.

Since all prioritarian social welfare functions satisfy these five axioms, this also holds for the two main SWF families used in this volume, namely those using the Atkinson and the Kolm-Pollak transformation functions. Let $g\left(w_{ij}\right) = \frac{w_{ij}^{1-\gamma}}{1-\gamma}$, where $\gamma > 0$ indicates the degree of inequality aversion (or

[16]More formally, if K < N people have the same levels of well-being in outcomes x and y, then the ranking between x and y depends only on the well-being levels of the other N-K people. This holds for any K < N.

priority for the worse-off), in Equation (3), and we have an Atkinson SWF. If, instead, we let $g(w_{ij}) = -e^{-\beta w_{ij}}$, with $\beta > 0$, we would have a Kolm-Pollak SWF, where β captures the degree of priority for the worse-off. These are well-established families of welfare functions, which can productively be applied to inform policy choices in a number of different domains.

It is perfectly possible, however, for two different social welfare functions satisfying the above axioms to rank two distributions, *x* and *y*, in opposite ways. Even within a given SWF family, it is perfectly possible that one member of the family (say, an Atkinson function with $\gamma = 1$) will rank *x* as preferable to *y*, whereas another (say, an Atkinson function with $\gamma = 4$) will rank *y* as preferable to *x*. The same can be said of different members of the Kolm-Pollak family, and indeed of other functional forms satisfying Equation (3). This is not a problem, per se. As described in Adler (2018), an observer or policymaker follows her own process of ethical deliberation in choosing a SWF, and that includes the choice of the inequality aversion parameter (e.g. γ, β). Given those preferences, it is perfectly appropriate for one such observer to prefer *x* to *y*, while another prefers *y* to *x.*

Yet the fact that rankings are in general dependent on specific choices of functional forms or parameter values is somewhat problematic when the objective at hand is to investigate whether prioritarianism as a broad approach to social justice is consistent or inconsistent with equality of opportunity as another broad approach to social justice. This is why we follow the robust rankings approach described in the Introduction: we search for conditions (about the relationship between the two distributions, ***w(x)*** and ***w(y)***) under which all SWFs in a given family will rank *x* and *y* the same way. As is standard in the welfare economics literature, this is achieved through theorems that establish the mathematical equivalence between (i) a unanimous ranking among all members of a family of social welfare functions and (ii) a dominance condition, typically expressed in terms of cumulative distributions functions, that can be tested empirically.

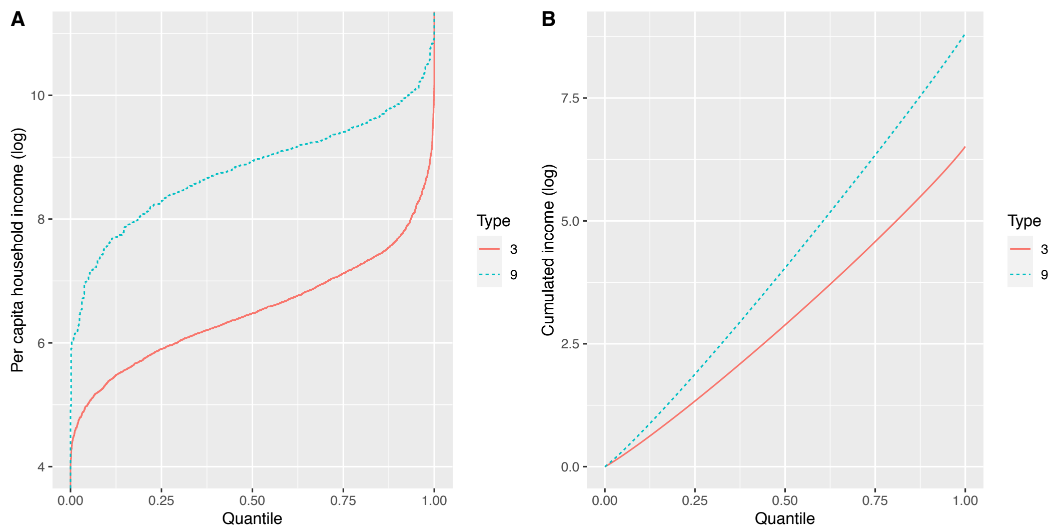
Two equivalence results are particularly interesting as they characterize dominance conditions that will be frequently used in this paper. The first is the equivalence between unanimous ranking in the broad family of SWFs defined by axioms I, II, IV and V above, without imposing the Pigou-Dalton transfer axiom (III), on the one hand, and first-order stochastic dominance (FOSD) on the other. A distribution ***w(x)*** is said to first-order stochastically dominate another, ***w(y),*** if at each and every rank, the element in ***w(x)*** is greater (or at least no less) than the corresponding element in ***w(y).*** That is, for all *k*, $w_k(x) \geq w_k(y)$. If this relationship holds between outcomes *x* and *y,* then Saposnik (1981) has shown that all SWFs in that broad family (which encompasses all prioritarian, as well as utilitarian and even various inequality-loving SWFs) will unambiguously rank *x* above *y*. Figure 2.A illustrates such a dominance result for two different types in South Africa in 2008.[17] Each of the two curves plots $w_k(x)$ as a function of k, for the respective type. Such curves are known as quantile functions and, as statistically minded readers will know, they are inverse functions of the cumulative distribution function (c.d.f.) of a distribution. As noted in Appendix A1, FOSD can also be – and indeed typically is – defined in terms of relationships between c.d.f.s.

Even more important for our purpose is the equivalence between unanimous rankings in the family of Prioritarian SWFs defined by axioms I – V above (now including axiom III: Pigou-Dalton) and second-order stochastic dominance (SOSD). A distribution ***w(x)*** is said to second-order stochastically dominate another, ***w(y)*** if, at each and every rank, the sum of all elements up until that rank is greater in ***w(x)*** than in ***w(y).***[18] If this relationship holds between outcomes *x* and *y,* then Shorrocks (1983) has shown that all prioritarian SWFs will unambiguously rank x above y. Figure 2.B illustrates such a dominance result for the same two South African type distributions shown in Panel A. In this case, each of the two curves plots cumulative well-being, $\frac{1}{N}\sum_{l=1}^{k} w_l(x)$ as a function of k, for the respective type (with income levels proxying for well-being). These curves are known as Generalized Lorenz Curves, written as $GL(k,x,N) = \frac{1}{N}\sum_{l=1}^{k} w_l(x)\,, k = 1, \dots, N$. To simplify notation, here we will denote the Generalized Lorenz Curve for any particular distribution *v(x)* simply as $GL(v(x),N)$, omitting the fact that the curve is a function of the rank k. When a distribution *v(x)* displays SOSD (or Generalized Lorenz Dominance) over *v(y),* we write $GL(v(x),N) > GL(v(y),N)$. Appendix A.1 contains the formal definitions of FOSD and SOSD, as well as a more formal statement of these two equivalence results.

[17] The empirical significance of the result is discussed in Section 5. The purpose of Figure 2 is merely to illustrate the concept.

[18] in general, this result is about the partial mean up to each rank: that is the sum is divided by N. In our setting, N is fixed and finite, so the result can be expressed as above.

Figure 2: Illustrations of First- and Second-Order Stochastic Dominance



Source: NIDS Wave 1. Note: Panel A displays the quantile functions for two groups of people in South Africa. The curve lying everywhere above the other is said to first-order stochastically dominate it. Panel B displays Generalized Lorenz curves for the same two groups. The curve lying everywhere above the other is said to second-order stochastically dominate it.

Equality of Opportunity

The above discussion of prioritarianism made no reference to people's efforts or circumstances. It could have been written without using any *i* or *j* subscripts: all that mattered for assessing social welfare in an outcome *x* was the distribution of individual well-being, $\boldsymbol{w}(x)$. How that distribution is generated and whether differences in well-being are due to differences in people's responsibility or effort choices (on the one hand), or merely to how rich their parents were (on the other), mattered not in the least.

That is not true in the normative theory of equality of opportunity (E. Op.), mentioned in the Introduction (and footnote 4). Under E.Op., whether differences in well-being are due to differences in a person's circumstances or efforts – i.e. to differences between rows or columns of Figure 1 – matters a great deal.[19] That is why a vector – $\boldsymbol{w}(x)$ – no longer contains all the information sufficient for ranking

[19] Because of the richer basal space in their approach, proponents of E.Op. see the theory as inherently non-welfarist and often describe it without using the words "well-being". Authors such as Roemer (1998) will speak of an "advantage" variable as the one that should be represented in Figure 1. An advantage has to be something that people value without strong satiation: more of it is never a bad thing. Empirical applications have used incomes,

social outcomes. A matrix $[W_{ij}]$ is needed. There are different versions of the theory, but all share two key principles, namely:

A. Principle of Compensation: Differences in "circumstances" beyond the control of the individuals warrant compensation, as they generate unfair inequalities in well-being.

B. Principle of Reward: "Efforts" should be rewarded, and the resulting inequalities in well-being should be preserved.

Although they can be stated in deceptively simple terms, and at first glance appear consistent with each other, these two principles can be defined in different ways and, in some cases, they are not actually mutually compatible. A substantial literature has now explored these differences and we will not dwell much on them here. The reader is referred to Fleurbaey and Peragine (2013) for some original results and to Ferreira and Peragine (2016) for a more general discussion. For our purposes it will suffice to distinguish between two different interpretations of the principle of compensation, and three alternative formulations of the reward principle.

In its *ex-ante* version, the principle of compensation is about compensating for inequalities between types (the rows in $[W_{ij}]$). Specifically, the approach requires (i) defining the opportunity set faced by each type, and (ii) specifying a manner for evaluating those sets. In most empirical applications, the *i*[th] row of the matrix in Figure 1, which is the support of the well-being distribution fo r type T_i, is used to represent the opportunity set of that type. Naturally, any number of summary statistics could be used to summarize each such vector, thereby evaluating the opportunity set: its mode or median for example. Historically, most authors have used the mean, $\mu(T_i) = \sum_j p_{ij} w_{ij} / \sum_j p_{ij}$.

The *ex-post* version of compensation, on the other hand, looks at inequality in well-being between individuals at each and every level of effort.[20] In other words, it focuses on inequality within tranches (the columns in $[W_{ij}]$). These two versions of the compensation principle have implications that are quite different. In fact, as established by Fleurbaey and Peragine (2013), the two versions are mutually

---

consumption expenditures, and measures of educational or health status as examples of advantage. Here we use "well-being" as the advantage in order to more easily integrate the discussion of E.Op. and prioritarianism. Furthermore, we remain agnostic as to the specific concept of "well-being": whether hedonic, preference-based, objective-good, etc...

[20] In the E.Op. literature, it has become established for the terms *ex-ante* and *ex-post* to be used to refer to "before" and "after" the realization of effort levels, rather than to before and after the resolution of uncertainty as, for example, in ex-ante and ex-post prioritarianism.

inconsistent in general, meaning that the policies (φ) chosen under one version might not be the same as those chosen under the other.

The reward principle is concerned with differences in well-being due to the exercise of personal responsibility. Since all individuals in each type share identical circumstances, the only thing that differs among them are the effort (or responsibility) levels e_j, so a reward principle is essentially about how to apportion different levels of well-being to different levels of effort, and so how to evaluate inequality in well-being within types and between tranches. This can be understood in a number of slightly different ways, of which we consider the following three.

The so-called *utilitarian reward* principle requires neutrality with respect to inequality in well-being within types: existing inequalities in well-being among individuals in the same type are a matter of social indifference. Alternatively, one could impose some degree of inequality aversion even within types – presumably to a lesser degree than between types. This view would correspond to a version of E.Op. where all inequalities are objectionable, but those due to differences in responsibility are less so. This is referred to as *inequality-averse reward*. Many different arguments can be used to motivate it. One practical argument is that, in empirical analysis, the researcher or policymaker can never expect to fully observe all elements of the vector of circumstances. To the extent that some circumstances remain unobserved, they "contaminate" $e(x)$, justifying some aversion to effort-driven inequality.[21] Finally, one can simply be agnostic about the degree of inequality aversion that should apply within types, perhaps because of the uncertainty just discussed. This approach is called, self-evidently, *agnostic reward*.

While these two versions of the compensation principle and three versions of the reward principle do not exhaust the variants proposed in the literature, they are varied enough to provide a solid basis for our discussion of whether E.Op. may be made consistent with simple prioritarianism, of the form given by Equation (3). If we restrict our attention to them, we can represent the variants of E.Op theory by a 2x3 matrix such as the one in Table 1.

[21] It has been argued – against the E.Op. approach – that all differences are ultimately driven by circumstances, genetic, social or otherwise. In that view, inequality aversion within and between types becomes identical, and we are back to pure prioritarianism. In principle, the degree of aversion to inequality within types might reflect the relative weight one places on "unobserved circumstances" relative to "free will" as driving the within-type differences in well-being.

Table 1: Variants of E.Op theory

	Utilitarian Reward	Inequality-averse Reward	Agnostic Reward
Ex-ante Compensation			
Ex-post Compensation			

The next two sections explore the nature of the clash between Prioritarianism and E.Op. under each of these six possible versions of the Equality of Opportunity approach. Section 3 considers the ex-ante approach to compensation: the first row of the matrix in Table 1. Section 4 turns to the bottom row and looks at the ex-post approach. We first show that Prioritarianism (as defined by the five axioms I-V above) is *in general* inconsistent with E.Op.. Then for each cell in Table 1, we propose adjustments to two of the axioms of prioritarianism, namely Anonymity and Pigou-Dalton, that can accommodate the relevant E.Op. principles. Adjustments are also needed to strengthen Separability, although these are more technical in nature and less substantive from a normative point of view. Under the modified – and more restrictive – social welfare functions implied by the new axioms, we show that a compromise between E.Op. and Prioritarianism *is* possible. Finally, drawing on earlier results in the literature, we state dominance conditions – variants of the first and second order stochastic dominance conditions discussed above – for classes of social welfare functions defined by the revised axioms. In Section 5, we illustrate these dominance results empirically for South Africa, demonstrating the existence of these compromises between prioritarianism and E.Op. in a real-world context.

3. Prioritarianism and equality of opportunity: the ex-ante case

As noted above, the ex-ante version of the compensation principle requires an evaluation of the opportunity set faced by each type, which is used to rank types. In practice, this is often accomplished by relying on a summary statistic (such as the type mean) to represent the value of the opportunity set of each type. The principle then requires policies to reduce the inequality among these values.

Returning to the matrix $[W_{ij}]$, represented by Figure 1, let us order the types such that mean well-being rises as we move down the table.[22] Recall that effort is defined so that well-being rises within each type as we move to the right along each row of Figure 1.[23] Then in general it is clearly possible that there exist two individuals, A and B, such that A is worse-off than B even though A belongs to a better-off type – that is to a type with a higher mean.[24] B is better-off than A because the rewards to her greater effort or responsibility more than compensate for the fact that she belongs to a lower-ranked type.

But therein lies a clash: According to the Pigou-Dalton Transfer Principle, if an outcome *y* (that is: a given matrix of well-being levels) is obtained from another outcome *x* exclusively by means of a transfer from A to B, with all other entries unchanged, then *x* should be preferred to *y*. This is a regressive transfer, from a poorer person (A) to a richer one (B). However, B belongs to a "poorer" (i.e. lower-ranked) *type* than A. The ex-ante principle of compensation therefore requires that we prefer outcome *y* to *x*, since inequality of opportunity (i.e. well-being differences due to circumstances) is lower in *y.* Inequality among type means is lower in *y*. This situation is represented in Figure 3: the transfer from A to B is regressive in terms of individuals, but progressive in terms of types. The Pigou-Dalton axiom of prioritarianism and the ex-ante compensation principle of E.Op. therefore clash. As stated so far, these two normative views are inconsistent: not all prioritarian social welfare functions S^P are consistent with the ex-ante principle of compensation.

[22] That is: $\mu(T_{i+1}) \geq \mu(T_i), i \in [1, \dots, n-1]$. In this discussion we use the type's mean to represent the value of each individual's opportunity set to simplify the presentation, but the conclusions would hold for any other scalar valuation of the sets.

[23] That is: $w_{ij} \leq w_{ij+1}, j \in [1, \dots, m-1]$. See also footnote 12.

[24] A has well-being w_{ij}, B has well-being $w_{i-b,j+k}$. It is possible that $w_{ij} < w_{i-b,j+k}, b > 0, k > 0$.

Figure 3: A stylized representation of the ex-ante EOp – prioritarianism clash

	e_1	e_2	e_3	…	e_m
C_1	w_{11}	w_{12}	w_{13}	…	w_{1m}
C_2	w_{21}	w_{22}	w_{23}	$w_{i-b,j+k}$ (B)	w_{2m}
C_3	w_{31}	w_{32}	w_{33}	…	w_{3m}
…	…	w_{ij} (A)	…	…	…
C_n	w_{n1}	w_{n2}	w_{n3}	…	w_{nm}

When, as in this case, two desirable properties or principles clash, compromises can be sought by restricting the domains of one or more of them. We therefore ask: are there sub-classes of social welfare functions, that are prioritarian "in spirit" but satisfy slightly different axioms, which might be consistent with ex-ante E.Op.? It turns out that the answer is yes. Consider replacing the Anonymity Axiom (I) from Section 2, with two different versions. The first is a partial symmetry property that applies only within types, while the second requires symmetry *of* (or among) types:

IA. Anonymity (or symmetry) within types: If an outcome *y* is obtained from another outcome *x* merely by re-arranging well-being levels w_{ij} among individuals *within types*, (so that a type distribution $\mathbf{w}_i(y)$ is a permutation of $\mathbf{w}_i(x)$, for all i), then one should be indifferent between *x* and *y*.

IB. Anonymity (or symmetry) of types: If an outcome *y* is obtained from another outcome *x* merely by re-arranging types (without changing the well-being levels w_{ij} of individuals *within types)*, then one should be indifferent between *x* and *y*.

Furthermore, similarly restrict the domain of the Pigou-Dalton transfer principle, as follows:

IIIA. Pigou-Dalton Transfer Between Types: Consider two types, *i* and *k*, with different means. If outcome *y* is obtained from outcome *x* by means of a finite sequence of pure (non-leaky) transfers exclusively between individuals in these two types (leaving all individuals in all other types unaffected), the net effect of which is that the richer type becomes even richer and the poorer type even poorer in *y*, then *x* is preferred to *y*.

These two new axioms merely re-define anonymity and Pigou-Dalton so that they hold not over the full domain of the distribution of well-being, but in more restricted domains. Axiom IA, for example, implies that swapping the well-being levels of Paul and Peter (while everyone else's well-being is unchanged, as in our earlier example) must leave aggregate social welfare unchanged if Paul and Peter share the same circumstances, but *may not leave it unchanged if they do not*. Axiom IB implies that social preferences depend not on the specific identity of each type (e.g. "Black men with highly educated parents", or "Asian women with parents with low education"), but on the type's relative rank. Axiom IIIA requires that a net transfer from a lower-ranked type to a higher-ranked type must lower social welfare, regardless of whether the particular individuals making the transfer are better or worse-off than those receiving it. Importantly, the axiom says nothing about transfers within a type.

Finally, replace the separability axiom (IV) with a two-part *Additivity Axiom*, which explicitly imposes an additive aggregation of well-being both within and between types:

IVA: Additivity

(i) *Between Types*: the social value of an outcome is equal to the sum of some (type specific) function of the well-being of each type.

(ii) *Within Types*: the well-being of a type is equal to the sum of some (individual specific) function of the well-being of individuals in that type.

It turns out that replacing Axioms I, III and IV from the characterization of a prioritarian social welfare function in Section 2 with Axioms IA, IB, IIIA and IVA can be used to define different classes of social welfare functions which *are* consistent with the ex-ante compensation principle.

To make further progress in specifying these classes and in stating the corresponding dominance conditions, we must distinguish among the three kinds of reward principle discussed earlier. Since this principle is concerned with rewarding effort or responsibility, and that is the only thing that differs among individuals within any given type, it is quite intuitive that one's attitude to reward is just a mirror image of one's attitude to inequality *within* types. If one is completely neutral with respect to inequality within types, then that implies that one subscribes to the utilitarian version of the reward principle. If one is still averse to inequality within types, although perhaps to a different extent than to inequality between types, then one subscribes to inequality-averse reward. Finally, one may decline to take either one of those ethical positions and choose to remain agnostic about inequality within-types, a view corresponding to the agnostic reward principle. In this latter case, one does not impose any specific axiom.

These give rise to three alternative versions of the reward principle, governing aversion to inequality within types, as follows:

VIA: *Inequality Neutrality within Types:* Suppose well-being levels are the same across outcomes *x* and *y* for all but two people, A and B, both of whom belong to type *i*, for any i , and respectively to tranches *j* and *k, j>k*. Suppose, furthermore, that y is obtained from x by means of a transfer from the poorer individual (B) to the richer one (A).[25] Then one is indifferent between *x* and *y*.

VIB: *Inequality Aversion within Types:* Suppose well-being levels are the same across outcomes x and y for all but two people, A and B, both of whom belong to type i, any i, and respectively to tranche *j* and *k*, *j<k*. Suppose that – as before – y is obtained from x by means of a transfer from the poorer individual (A) to the richer one (B). Then *x* is preferred to *y*.

VIC: *Inequality Agnosticism within Types:* No axiom is imposed.

Combining Axioms IA, IB, II, IIIA, IVA, V, and VIA yields a class of welfare functions given by:

$$S^N = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} g_i(w_{ij}) \tag{4}$$

where the functions $g_i(w_{ij})$, which may now vary across types, are all linear and become progressively less steep as types get “richer”, or better-off.[26] (Peragine, 2004).

In other words, consider a person whose views on social justice are such that they value an increment in well-being for any person positively if no one else loses (Monotonicity); they believe well-being should be aggregated additively both within and across types (Additivity); they believe that people who share the same set of circumstances should be treated impartially (Anonymity within types), just as types themselves should also be treated symmetrically (Anonymity between types: what matters is a type’s income vector, which determines its rank; not its “identity”); and that compensation should be

[25] That is: $w_{ij}^A(x) > w_{ik}^B(x)$ and $w_{ij}^A(y) = w_{ij}^A(x) + \delta$ and $w_{ik}^B(y) = w_{ik}^B(x) - \delta$.

[26] More formally, $g_i(w)$ satisfy three properties. Within any type T_i, $g_i'(w) > 0$, $g_i''(w) = 0$. And $g_i'(w) > g_{i+1}'(w) > 0, \forall i \in [1, \dots, n-1]$. Recall that types are ordered by their means; see footnote 22.

made for inequality between types but not for any inequality within types (Pigou-Dalton between types and Inequality neutrality within types). Such a person would rank outcomes in ways consistent with a SWF given by (4). Let's call all such people "Strict Opportunity Egalitarians", or SOEs.

Three key properties of (4) – in contrast to (3) – are worth noting. First, the transformation function $g(w)$ is now type-specific, rather than being identical for everyone in society. Second, each such type-specific transformation function is linear – and thus insensitive to redistribution within types. But third, its slope is lower as types become richer, meaning that transfers from a richer type to a poorer type increase overall social welfare. A specific – and quite intuitive – member of S^N would be a weighted sum of type means, where the weights decline with the type mean.[27] This would be the discrete-setting equivalent of a sum (across types) of "concave" transformations of type means.

Having defined our first class of "opportunity-prioritarian" SWFs, we are now ready to state our first equivalence result. To do so, let's define the distribution of type well-being, $M(x)$, as the *n x 1* vector the elements of which are the sums of well-being levels accruing to all individuals in each type, in outcome *x*.[28] An important result due to Peragine (2004) is that if (and only if) a person's ethical views can be represented by <u>any</u> social welfare function in the S^Nclass, then this person will always prefer an outcome x over another outcome y if the distribution of total type well-being in *x* displays Generalized Lorenz Dominance over that in y, that is:

$$S^N(x) > S^N(y) \iff GL(M(x), n) > GL(M(y), n) \qquad (5)$$

This statement is analogous to Shorrocks's (1983) Theorem reproduced in Appendix A.1. That theorem established that if a distribution of <u>individual</u> well-being, $\boldsymbol{w}(x)$, Generalized Lorenz dominated another, $\boldsymbol{w}(y)$, then all prioritarian social welfare functions would rank *x* higher than *y* (and vice-versa). Peragine's (2004) Theorem 1 says that if a distribution of <u>type</u> well-being $M(x)$ Generalized Lorenz dominates another, $M(y)$, then all SOEs will rank x higher than *y* (and vice-versa). This gives us a result to fill in the first cell in Figure 2: a comparison of Prioritarianism and EOp when the latter adopts the ex-ante version of the compensation principle and the utilitarian version of the reward principle.[29]

[27] For example: $g_i(w_{ij}) = n_i^{-1} \alpha^{n+1-i} \sum_{j|i} w_{ij}, \alpha > 1$.
[28] That is, each element of $M(x)$ is given by: $M_i = \sum_{j=1}^{m} p_{ij} w_{ij}(x)$.
[29] It is worth noting that if one takes aversion to inequality between types to an extreme, then $g_1{}'(w) > 0$ and $g_i'(w) \to 0, \forall i > 1$. Then all that matters is the poorest type, and the dominance conditions on $M(x)$ collapse to the first element of that vector of total incomes: $n_1 \mu(T_1, x) > n_1 \mu(T_1, y)$. This is the partial ordering equivalent to

If, instead, one subscribes to the inequality-averse reward principle, then it is axiom VIB that one wants to combine with axioms IA, IB, II, IIIA, IVA and V. This combination yields a slightly different class of social welfare functions: it is still given by (4) and $g_i(w_{ij})$ continues to vary across types and to become progressively less steep as types get "richer," or better-off. But with axiom VIB, the transformation functions within each type are no longer linear; they are strictly concave instead.[30] We call this class S^V, and those who subscribe to it the "Inequality-Averse Opportunity Egalitarians" or IOEs. The only difference between S^V and S^N is, of course, that the second derivative of the type-specific transformation function is negative: $g_i''(w) < 0$. This imposes inequality-aversion within types.

Now let $n_i = \sum_{j=1}^{m} p_{ij}$ denote the population of type T_i and the type-specific Generalized Lorenz Curve for T_i as $GL_i(w_i(x), n_i)$.[31] It turns out that a theorem due to Atkinson and Bourguignon (1987), which was stated and proved in a different context, can be straightforwardly re-interpreted to establish the dominance conditions analogous to (5) which apply for social welfare functions in S^V. The theorem establishes that:

$$S^V(x) > S^V(y) \iff \sum_{i=1}^{k} GL_i(w_i(x), n_i) \geq \sum_{i=1}^{k} GL_i(w_i(x), n_i)\,, k = 1, \dots, n \qquad (6)$$

In other words: Shorrocks's (1983) Theorem stated that Generalized Lorenz dominance in the full distribution (across the entire population) of *x* over *y* implied that all prioritarian Social Welfare Functions (i.e. all members of the S^P class) would rank x as preferred to y. Peragine's Theorem in (5) told us that, if instead of being a Prioritarian, you were a Strict Opportunity Egalitarian, whose welfare functions were of the form S^N, then the relevant condition was Generalized Lorenz dominance not of the full distribution of individuals, but of the type's well-being distribution. Now we learn from the Atkinson-Bourguignon (1987) Theorem, re-stated in (6), that if you reject inequality neutrality within types (that is, utilitarian reward), but require instead some degree of inequality-aversion within types, then the relevant condition is sequentially additive second-order dominance of the type-specific Generalized Lorenz curves.

---

van de Gaer's allocation rule under the ex-ante compensation principle of E.Op., namely, to maximize the mean well-being of the poorest type – but here those means and population weighted.

[30] In this case within any type T_i, $g_i'(w) > 0$, $g_i''(w) < 0$. Although strict concavity is now permitted, it must be the case that $\inf g_i'(w) > \sup g_{i+1}'(w) > 0, \forall i \in [1, \dots, n-1]$, so as to avoid a potential clash with Axiom IIIA.

[31] Specifically, $GL_i(w_i(x), n_i) = \frac{1}{n_i}\sum_{j=1}^{q} w_{ij}(x)$.

A similar result exists also for the combination of ex-ante compensation and agnostic reward. Here we combine axioms IA, IB, II, IIIA, IVA and V only, without imposing any reward axiom; that is without imposing any restrictions on our attitude to inequality within types. This generates a class of social welfare functions still given by (4), but where the functions $g_i(w_{ij})$ satisfy only two properties: (i) Within any type T_i, $g_i'(w) > 0$. And across types, (ii) $\inf g_i'(w) > \sup g_{i+1}'(w) > 0, for\ all\ i \in [1, \dots, n-1]$. We call this class S^A, and those who hold such preferences "Agnostic Opportunity Egalitarians" (AOEs). The difference between this class and the two considered earlier is the absence of any restriction on the second derivative of the type-specific transformation functions $g_i(w_{ij})$: Agnosticism with respect to inequality within-types means that the transformation function within each type has to be upward-sloping, but it can be strictly concave, linear, or even strictly convex. Naturally, this yields a larger set of welfare functions, of which the two disjoint sets of S^N and S^V are both strict subsets. The dominance criterion for S^A will therefore be correspondingly stronger.

Indeed, Peragine and Serlenga (2008) find that a unanimous ranking of outcome *x* over outcome *y* among all social welfare function in the S^A class requires a sequential first-order stochastic dominance of population-weighted type-distributions between outcomes *x* and *y:*

$$S^A(x) > S^A(y) \iff \sum_{i=1}^{k} n_i\, F_i\big(w_{ij}(x)|T_i\big) < \sum_{i=1}^{k} n_i\, F_i\big(w_{ij}(y)|T_i\big), for\ k = 1, \dots, n \tag{7}$$

In (7), $F_i\big(w_{ij}(x)|T_i\big)$ denotes the (discrete) cumulative distribution function of well-being in Type T_i, under outcome *x*.[32] The algorithm to check dominance is as follows: one first compares the distributions of the poorest type, in outcomes *x* and *y*. The c.d.f. in x must lie everywhere below that in y.[33] Then one adds the distribution of the second-poorest type, thereby creating a mixture of the distributions of the two lowest types under each outcome. Again, that mixture in x must first-order dominate y. Carry on creating mixtures of distributions by sequentially bringing in the next poorest type. If first-order-dominance holds *n* times, then all SWFs of the S^A form will rank x higher than y.

The foregoing discussion allows us to fill in the Axioms and Results for the first row of Table 1, which we do below in Table 2:

[32] In general, we use the mathematical notation $(a|b)$ to denote "a conditional on b".
[33] More precisely, the c.d.f. in x must lie nowhere above and at least somewhere below that in y.

Table 2 Ex-ante compensation: axioms and results

	Utilitarian Reward	Inequality-averse Reward	Agnostic Reward
Ex-ante Compen sation	**Axioms:** Anonymity within Types; Anonymity of Types; Monotonicity; P-D Between Types; Additivity; Continuity; Inequality Neutrality within Types. **Dominance condition:** Generalized Lorenz Dominance of the total type well-being distribution	**Axioms:** Anonymity within Types; Anonymity of Types; Monotonicity; P-D Between Types; Additivity; Continuity; Inequality Aversion within Types. **Dominance condition:** Sequential second-order dominance of the type-specific Generalized Lorenz curves.	**Axioms:** Anonymity within Types; Anonymity of Types; Monotonicity; P-D Between Types; Additivity; Continuity. **Dominance condition:** Sequential first-order dominance of population-weighted type-distributions
Ex-post Compen sation			

Discussion

There is a clear analogy between these three dominance results and their well-known antecedents in welfare economics. For outcome *x* to be preferred to outcome *y* according to *any* SWF that satisfies anonymity, monotonicity, continuity and separability (but not necessarily Pigou-Dalton), one needs first-order dominance of the distribution in *x* over *y*. If we add an inequality-aversion requirement – the Pigou-Dalton axiom – this makes the SWFs not only increasing but also strictly concave. That is a smaller set of functions, so it is "easier" to obtain dominance: instead of first-order stochastic dominance, only second-order is needed. That is, instead of the cumulative distribution function in x lying everywhere below that for y, we need the Generalized Lorenz for x lying everywhere above that for y. And if we are strictly neutral with respect to inequality, then a comparison of means or sums is sufficient to rank *x* and *y*.

When we try to make social welfare rankings of that kind consistent with the two fundamental principles of equality of opportunity (compensation and reward), we face a clash: those two principles are generally inconsistent with two of the axioms used earlier in welfare rankings: anonymity and Pigou-Dalton. As we have seen in this section, it is possible to reconcile the two principles with social welfare functions, but only if one is prepared to limit their domains. In the ex-ante case that we have seen so far, anonymity must apply only partially: among people in the same type, and among the types themselves –

but not among all people. Pigou-Dalton must apply only between types. With those restrictions, a compromise between welfarism (Prioritarianism) and E.Op. can be obtained. Its exact nature depends on one's specific views on the reward principle.

If we choose to take the notion that all inequality within types is ethically acceptable and not to be compensated – that is: if we are Strict Opportunity Egalitarians – then we are neutral to inequality within types, and we can define a set of SWFs (those with the form and properties of W_N), all members of which will agree in their rankings across outcomes *x* and *y* if and only if we observe Generalized Lorenz Dominance of the total type well-being distributions between x and y. The total type well-being distribution is just a vector of the total well-being accruing to each type, for all types. Total well-being is, of course, mean well-being in a type multiplied by the type's population.

If, on the other hand, we are not quite so convinced that all inequality within-types should be accepted, then we are averse to inequality within types. Agreement among all SWFs with the form and properties of S^V will obtain if and only if a sequence of Generalized Dominances is observed for the type distributions (rather than just their means). This more demanding condition reflects the inequality aversion within types: if we don't care about this distribution within types, all that matters are the type means, or total sums. But if we do, we need the kind of second-order dominance we saw in Shorrocks's Theorem, but now applied to each type (and then aggregated sequentially across types, from the lowest to the highest-ranked).

As always, dominance conditions become more demanding if we are seeking agreement across a larger set of SWFs. So: if we insist on being completely agnostic about inequality within types – allowing that second derivative to be anything from negative to positive infinity – then sequential second-order dominance of type distribution is no longer enough: first-order dominance is needed, for each step in the sequential summation of types (mixtures of distributions), from the poorest on upwards, until you have the full distribution.

We now turn to a discussion of the clash between Prioritarianism and EOp – and of possible axiomatic compromises to reconcile them – under an ex-post version of the compensation principle.

4. Prioritarianism and equality of opportunity: the ex-post case

The ex-post compensation principle requires reducing the inequality in well-being among individuals exerting the same level of effort. The literature has proposed different versions of this principle and, in addition, different social criteria have been formulated which combine ex-post compensation and the different versions of the reward principle. In this section we briefly review this literature, focusing on the contributions that have adopted a social welfare approach to rank distributions. In general, much as for the ex-ante case discussed in the previous section, the features that distinguish the opportunity-egalitarian social welfare orderings from the prioritarian are the social evaluations of inequality-reducing operations (such as the Pigou Dalton transfer) and of permutations of individuals (expressed by the symmetry axiom).

While in the prioritarian approach any inequality-reducing transfer has a positive impact on social welfare, in the ex-post opportunity-egalitarian approach the social welfare judgment on such operations will be conditioned on the effort level exerted by the individuals involved in the transfer (the tranche). This is in contrast to the ex-ante approach, where such judgment is conditioned on the circumstances (the type) of the individuals involved. Similarly, while in the prioritarian approach no permutation of individuals has any effect on social welfare – meaning precisely that only individual well-being matters for social welfare – in the ex-post opportunity egalitarian approach the social welfare judgment on permutations will once again be conditioned on the effort level of the individuals involved. We now turn to the main criteria consistent with this view.

In his seminal contribution, Roemer (1993) proposes two different ex-post equality of opportunity criteria. According to the first criterion, the social objective should be to maximize the minimum welfare level for each degree of effort: that is to say, the social objective is the maximization of the minimum of each tranche distribution. This criterion clearly expresses extreme inequality aversion within tranches combined with an agnostic view on the inequality between tranches, i.e., an agnostic version of the reward principle. Roemer's original proposal was formulated in a context of optimal taxation, where the objective was to find the tax rate which maximizes the social welfare. Interpreted in the current context of social rankings, Roemer's first criterion can be defined as follows: one distribution *x* is preferred to another distribution *y* if, at each tranche *j*, the minimum in *x* is higher than the minimum in *y*. Formally:[34]

$$S(x) > S(y) \iff min\left(T^j(x)\right) > min\left(T^j(y)\right),\ for\ all\ j = 1, \dots, m \qquad (8)$$

[34] Equivalently: $S(x) > S(y) \iff \min_i\left[w_{ij}(x)\middle|e_j\right] > \min_i\left[w_{ij}(y)\middle|e_j\right], for\ all\ j = 1, \dots, m$

Peragine (2004) generalizes Roemer's first criterion by weakening the property of extreme inequality aversion within tranches and proposing instead a less extreme view of ex-post compensation, based on a revised version of the Pigou Dalton transfer principle. The property he proposes, *Within Tranches Pigou Dalton,* requires that any reduction of inequality among individuals in the same tranche should be welfare improving:

Within Tranches Pigou Dalton: Consider outcomes x and y, and assume they differ only for tranche *j*. If outcome *y* is obtained from outcome *x* by means of a finite sequence of pure (non-leaky) rank preserving progressive transfers between individuals in tranche j, then *y* is preferred to *x.*

This latter property is combined with the following properties:

- *Anonymity Within Tranches:* any permutation within tranches does not change the social welfare);
- *Additivity Between Tranches:* the social value of an outcome is equal to the sum of some (tranche specific) function of the well-being of each tranche;
- *Monotonicity:* any increment to individual income is welfare improving.

The axioms above yield the characterization of the following family of social welfare functions:

$$S^T(x) = \sum_{j=1}^{m} v_j\big(w_{ij}(x)\big|e_j\big) \tag{9}$$

where the functions v_j, which may vary across tranches, satisfy two properties: within any tranche T^j, $v_j' > 0,\ v_j'' < 0$ (Peragine, 2004). Hence, social welfare is given by an additive aggregation of the welfare across tranches, where each tranche is evaluated by an increasing, symmetric and strictly concave function. Although in this general form, (9) continues to impose no restrictions on inequality between tranches – that is, it remains agnostic about reward – one can easily conceive of specific members of S^T that would introduce a specific reward scheme. The v_j function might, for example, use a single strictly concave transformation function for all tranches, but multiply the sum of transformed well-being within each tranche by some factor increasing in effort.[35]

Based on the above family of social welfare functions Peragine (2004) obtains a suitable dominance condition, according to which an outcome x is preferred to another outcome y if and only if, at each

[35] For example, $v_j\big(w_{ij}(x)\big|e_j\big) = \alpha^j \sum_{i=1}^{n} g\big(w_{ij}(x)\big|e_j\big)$, $g' > 0, g'' < 0, \alpha > 1$

tranche, the distribution in x, $w_j(x)$, Generalized Lorenz dominates the distribution in y, $w_j(y)$. Formally, let $n_j = \sum_{i=1}^{n} p_{ij}$ denote the population of tranche T^j, and $GL_j\left(w_j(x), n_j\right)$ denote the tranche-specific Generalized Lorenz Curve for T^j*(x)*.[36] Peragine (2004)'s theorem establishes that:

$$S^T(x) > S^T(y) \iff GL_j\left(w_j(x), n_j\right) > GL_j\left(w_j(y), n_j\right), for\ all\ j = 1, \dots, m \tag{10}$$

This result is the "ex-post opportunity egalitarian" counterpart of Shorrocks (1983) characterization of the Generalized Lorenz dominance based on the family of (generalized) utilitarian social welfare functions.

Let us turn now to Roemer's (1993) second criterion. Keeping the extreme inequality aversion within tranches, Roemer's second proposed solution was to maximize the average of the minimal values of the tranche distributions. That is to say, the social planner should first identify the minimum of each tranche; then, she should take the mean of these minimal values. This quantity becomes the social objective: the "mean of (tranche) mins" criterion. This criterion endorses extreme inequality aversion within tranches and a utilitarian version of the reward principle. Formally:

$$S(x) > S(y) \iff \frac{1}{m}\sum_{i=1}^{m} min\left(T^j(x)\right) > \frac{1}{m}\sum_{i=1}^{m} min\left(T^j(y)\right) \tag{11}$$

Clearly this is a much less demanding criterion than Roemer's first condition: only one dominance test is required, instead of the *m* tests for Roemer's first condition. In a recent paper, Fleurbaey, Peragine and Ramos (2017 – henceforth FPR) generalize both of Roemer's criteria by weakening the extreme within-tranche inequality aversion and focusing on the concept of the envelope of the type distributions. As seen above, Roemer (1993) proposes to look only at the worst-off individuals in each tranche. FPR's contribution is close in spirit to Roemer's proposal. However, in their approach not only the worst off, but also the second worst off, and the third worst off and so on are taken into account. More precisely, they define a *class* as a set of individuals that sit at the same position in their respective tranche distributions. The first class is exactly Roemer's maximand, but FPR consider all classes. Members of the same class are in the same position in their respective tranche distribution, meaning that the impact of circumstances is similar for all of them, when this impact is evaluated by their well-being ranking in their tranche. The Fleurbaey et al. ex-post criteria are then based on the idea of reducing the inequality between classes.

[36] Specifically: $GL_j\left(w_j(x), n_j\right) = \frac{1}{n_j}\sum_{i=1}^{q} w_{ij}(x)$.

Their proposal works as follows. Starting from the original matrix, first rearrange the values within each tranche until all tranches contain rank ordered welfare levels, from the lowest to the highest. In this way, we obtain a transformed matrix, whose rows are now labeled "classes." [37] The focus now becomes the reduction of inequality between classes. The authors then impose the crucial property of *Pigou Dalton Transfer Between Classes*, which states that any progressive transfer between two classes improves social welfare:

> *Pigou Dalton Transfer Between Classes:* Consider two classes, *i* and *k*, with different means. If outcome *y* is obtained from outcome *x* by means of a finite sequence of pure (non-leaky) transfers exclusively between individuals in these two classes (leaving all individuals in all other classes unaffected), the net effect of which is that the richer class becomes even richer and the poorer class even poorer in *y*, then *x* is preferred to *y*.

This property, inspired by ex-post compensation, is then combined with *Monotonicity, Anonymity Within Classes (requiring that any permutation within a class leaves social welfare unchanged)*, and three different versions of the reward principle: *Utilitarian, Agnostic* and *Inequality Averse Reward* within classes. Correspondingly, they obtain three different dominance conditions, all expressed as sequential dominances of the class distributions.[38]

When Utilitarian Reward is used, a unanimous ranking $S(x) > S(y)$ requires Generalized Lorenz dominance of the distribution of Class means, defined analogously to the distribution of Type well-being discussed in Section 3 (see Eq. 5), but with the means taken over the rows of the re-organized Class matrix, as opposed to the original matrix where rows were types.[39] When Agnostic Reward is used instead, unanimous rankings require an additive sequence of first-order dominance among class distributions, analogous to that between types, described in Eq. (7). Finally, when inequality-averse reward within class is imposed, unanimity requires an additive sequence of Generalized Lorenz dominance among class distributions, analogous to that described in Eq. (6). Formal statements of these conditions are consigned to Appendix A.2.

As noted in footnote 37, if the type distributions are characterized by sequential first-order dominance – that is, if for each level of effort, being in a higher type implies having higher level of well-

[37] Note that, in general, classes are not the same as types because type distributions may cross: one type may "lie below" another over a given set of tranches, but above it for a different set of tranches. In the special case when there are no "type crossings", types and classes are the same.
[38] Note that Fleurbaey et al. (2017) do not impose *Continuity.*
[39] Recall, once again, that in our setup with N constant, these conditions can be expressed interchangeable in terms of sums or means.

being – then the classes correspond to the types. In this case, the ex-post dominance conditions characterized by FPR correspond to the ex-ante dominance conditions discussed in the previous section. More precisely, condition (i) in Appendix A.2 corresponds to the ex-ante dominance characterized by Peragine (2004); condition (ii) to the condition in Peragine and Serlenga (2008); and condition (iii) to the sequential generalized Lorenz dominance of Atkinson and Bourguignon (1987).

This review allows us to complete our comparison of axioms and results embodied in Tables 1 and 2, which we do in Table 3:

Table 3: Ex-ante and ex-post compensation: axioms and results

	Utilitarian Reward	Inequality-averse Reward	Agnostic Reward
Ex-ante Compensation	**Axioms:** Anonymity within Types; Anonymity of Types; Monotonicity; P-D Between Types; Additivity; Continuity; Inequality Neutrality within Types. **Dominance condition:** Generalized Lorenz Dominance of the total type well-being distribution	**Axioms:** Anonymity within Types; Anonymity of Types; Monotonicity; P-D Between Types; Additivity; Continuity; Inequality Aversion within Types. **Dominance condition:** Sequential second-order dominance of the type-specific Generalized Lorenz curves.	**Axioms:** Anonymity within Types; Anonymity of Types; Monotonicity; P-D Between Types; Additivity; Continuity. **Dominance condition:** Sequential first-order dominance of population-weighted type-distributions
Ex-post Compensation (using Classes)	**Axioms:** Anonymity within Classes; Monotonicity; P-D Between Classes; Inequality Neutrality within Classes. **Dominance condition:** Generalized Lorenz Dominance of the distribution of class means	**Axioms:** Anonymity within Classes; Monotonicity; P-D Between Classes; Inequality Aversion within Classes. **Dominance condition:** Sequential second-order dominance of the class-specific Generalized Lorenz curves.	**Axioms:** Anonymity within Classes; Monotonicity; P-D Between Classes. **Dominance condition:** Sequential first-order dominance of Class distributions
Ex post Compensation (using Tranches)			**Axioms:** Anonymity within Tranches; Monotonicity; P-D Within Tranches; Additivity Between Tranches; Continuity.

			Dominance condition: Generalized Lorenz Dominance within all tranches.

Just as in Section 3, footnote 29 noted that van de Gaer's "min of means" allocation rule corresponded to a special, extreme case of the combination of ex-ante compensation and utilitarian reward (Cell 1,1 in Table 3), the foregoing discussion in this section has indicated that Roemer's "mean of mins" allocation rule (described above as his second criterion) is an extreme case of the combination of ex-post compensation and utilitarian reward (Cell 2,1). In both cases, taking inequality aversion to an extreme, leads to an exclusive focus on the mean of the lowest-ranked type (in one case) or class (in the other). The use of the mean arises from the utilitarian version of the reward principle, which mandates inequality neutrality within the relevant group (type or class).

Analogously, a special case of Roemer's first criterion – which focuses on the full lower envelope of tranches – can be obtained by taking inequality-aversion between envelopes (i.e. classes) to an extreme in a combination of ex-post compensation and agnostic reward (Cell 2,3 in Table 3). Agnostic reward is extremely demanding, requiring first-order dominance among the relevant groups (classes, in this case). With extreme inequality aversion, all that matters is the lowest class, i.e. the lower envelope of tranches. Finally, in cell (3,3), Peragine (2004) shows that Generalized Lorenz dominance within each and every tranche yields the same result as sequential first-order dominance of class distributions – as should be intuitively clear.

The discussion in Sections 3 and 4, drawing on various existing results from the literature, has enabled us to summarize the kinds of desiderata (axioms) that can be imposed on social orderings (or welfare functions) to combine the central features of prioritarianism (inequality aversion) and E.Op. theory (compensation for circumstances; reward to effort). We first established that the standard axioms of prioritarianism, in their pure form, clash with the compensation principle – whether in its ex-ante or ex-post forms. In essence, this is because prioritarianism is a welfarist criterion: the distribution of well-being contains all the information needed to rank societies (outcomes). In contrast, the normative theory of equality of opportunity is non-welfarist: the distribution of well-being does *not* contain all the information needed to rank societies; additional information is needed.

However, then we went on to note that suitable restrictions on two key axioms underpinning prioritarianism – namely anonymity and the Pigou-Dalton transfer principle – could yield narrower classes of welfare functions that still satisfied (more limited versions of) impartiality and inequality aversion, while also allowing for differential roles for circumstances and efforts in assessing distributions of well-being.

The complexity, if any, arose from the fact that the principles of compensation and reward in the normative theory of equality of opportunity can be, and have been, formally defined in different ways. Table 3 summarizes seven such combinations of these principles which, while not exhaustive of the literature, cover a meaningful span of the normative choices one must make.

For each such combination of compensation and reward principle, we noted which axioms characterize the relevant classes of social orderings and described the dominance criteria between distributions that would be required to establish rankings that are robust to changes in functional form within each class. These requirements vary substantially: from comparing vectors of type or class means, to elaborate sequences of first-order dominance results. In practice, do these differences also imply widely different results – so that comparisons across distributions are highly sensitive to which particular combination of compensation and reward principles (i.e. which cell in Table 3) one happens to choose? Although it is impossible to answer this question comprehensively here, the next section sheds some light on it by turning to one empirical example: an application of the comparisons just outlined to a ten-year period in the recent history of South Africa, possibly the world's most unequal country today.

5. An empirical application: The case of South Africa

To investigate the various opportunity-prioritarian criteria for comparing social welfare that were proposed above, we use data from all waves of the South Africa National Income Dynamics Study (NIDS). There are five such waves, recording information from 2008, 2010/11, 2012, 2014/5 and 2017. NIDS was designed to follow the original sample of households over time. However, using suitable weights, each wave can also be treated as a cross-section survey representative of the South African population. Like most household surveys, the NIDS does not contain a measure that would truly correspond to the concept of "individual well-being," discussed in the foregoing sections. In common with most of the empirical literature, in this section we approximate well-being by monthly per capita household income. That variable includes all regular income received by the household on a monthly basis, net of taxes, as well as imputed rental income from owner-occupied housing.[40] Besides income, the NIDS also contains information about a number of circumstances beyond individual control, notably: parental education, parental occupation, ethnicity and area of birth.

In our analysis, we restrict the sample to adult individuals aged 18-65. Missing information further limits the sample that can be used to estimate distributions: for example, we exclude the circumstance "area of birth" because of the extremely high share of missing values (above 50% in four waves). We keep in the sample only observations with non-missing information for all the other circumstances, in at least one period.[41] Additional information about missing data is reported in Appendix A.3.

We exploit the longitudinal nature of the survey, which includes probability weights calibrated to keep the samples representative of the population interviewed in Wave 1 over time (Brophy et al., 2018). Panel weights are set to zero for households that entered the survey in later waves, including the 2017 top-up sample introduced to improve the representativeness of top incomes in the survey. As a consequence, our empirical exercise does not assess the evolution of well-being or inequality of opportunity in South Africa over time. What it does evaluate is the dynamics of social welfare for the population originally sampled in 2008.[42]

Descriptive statistics for the samples used in our analysis are reported in Table 4. Average disposable income fell between 2008 and 2010/11, but gradually rose thereafter. By 2017, mean per capita income was 17% above its initial 2008 level. Income inequality was extremely high by international standards during the whole period but declined markedly over the period; particularly after 2010.

[40] There is a large literature on the fraught and incomplete nature of income or consumption expenditure variables as proxies for well-being. The limitations are clearly important, but we do not discuss them further here.

[41] Following individuals over time, whenever information on a circumstance is missing in a given wave, but not missing in other waves, we impute the latest non-missing value.

[42] For the same reason and because the two re-rankings of types' mean income (in 2010/11 and 2014/5) are not statistically significant, the ranking of types is fixed in 2008 and subsequently preserved. The use of weights allows us to correct for attrition. Brophy et al., (2018) report and attrition rate be between 22% and 14%, depending on the waves, and involving especially White, Indian/Asian and high-income respondents.

Table 4: Descriptive statistics for the samples included in the analysis

Wave	Sample size	Age	Income	Gini
2008	7,353	39.09	3,467.75	0.6879
2010/11	4,815	40.66	3,224.56	0.6844
2012	5,896	41.03	3,591.67	0.6574
2014/15	6,443	40.81	3,747.47	0.6262
2017	4,397	43.37	4,065.30	0.6283

Source NIDS W1-W5. Note: monthly disposable per capita incomes are in RAND 2015.

To compare the various alternative opportunity-prioritarian criteria discussed above, we first need to partition the sample into types and tranches, as in Figure 1. Selecting a type-partition is not a trivial matter, since there is an empirical trade-off between comprehensiveness (seeking to limit the downward biases arising from the partial observability of circumstances) and precision (that arises from limiting the risk of overfitting the model, which can in principle bias estimates upward); see Brunori et al. (2019). We follow Brunori et al. (2018) and use a conditional inference regression tree to obtain our optimal type partition. This particular machine-learning algorithm has been shown to perform well when estimating inequality of opportunity on survey data (Hothorn et al., 2006; Brunori et al., 2018).[43]

Figure 4 shows the opportunity tree obtained for South Africa in 2008. Independent variables used to split the sample into types are the circumstances beyond individual control already mentioned: race (four categories), father's and mother's education (five ordinal values each), father's and mother's occupation

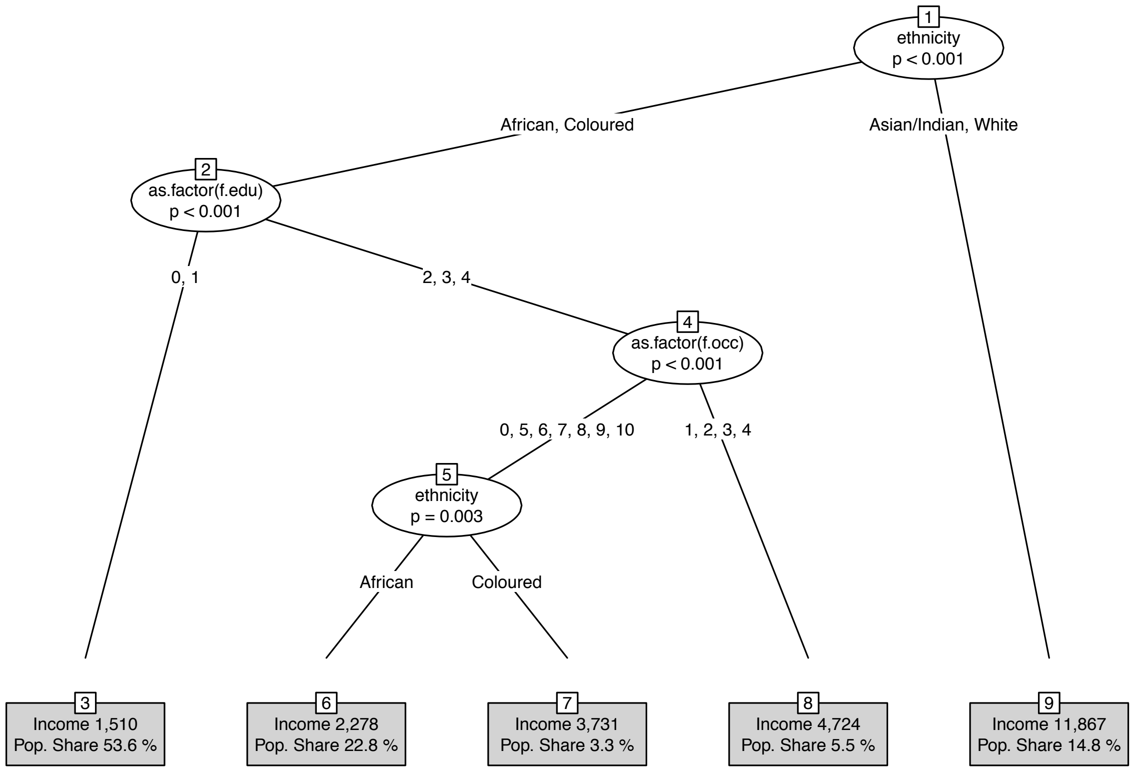
[43] In order to obtain a robust type partition we control the growth of the tree by setting two rather conservative requirements: first we set the confidence level (1- α) = 0.99; second we impose a minimum number of observations per type across all waves (100). The algorithm then obtains the partition in types iterating the following six steps:

1. The algorithm first tests the degree of correlation between the dependent variable (per capita income) and all the observable independent variables (circumstances);
2. If all tests have a Bonferroni-adjusted p-value higher than α the algorithm stops;
3. If one or more regressors have an adjusted p-value lower than α the algorithm selects the regressor with the lowest p-value;
4. Then the algorithm considers all values of such regressor as possible splitting points, that is the value used to partition the population into two subgroups. For each value and resulting subgroups, the algorithm tests whether the means of the dependent variable in the two subgroups are significantly different from each other;
5. The splitting value selected is again the value producing the lowest p-value of the test;
6. Steps 1 to 5 are repeated for all resulting subgroups.

The algorithm eventually stops at step 2 and the resulting partition can be represented as an upside-down tree.

(11 categories each). Alongside the name of the variable, each splitting point reports the Bonferroni-adjusted *p*-value of the correlation test. Terminal nodes describe the partition in five types and report expected incomes and population shares. Corresponding types' empirical cumulative distribution functions are shown in Figure 5 (the same distributions for subsequent waves are in Appendix 2).

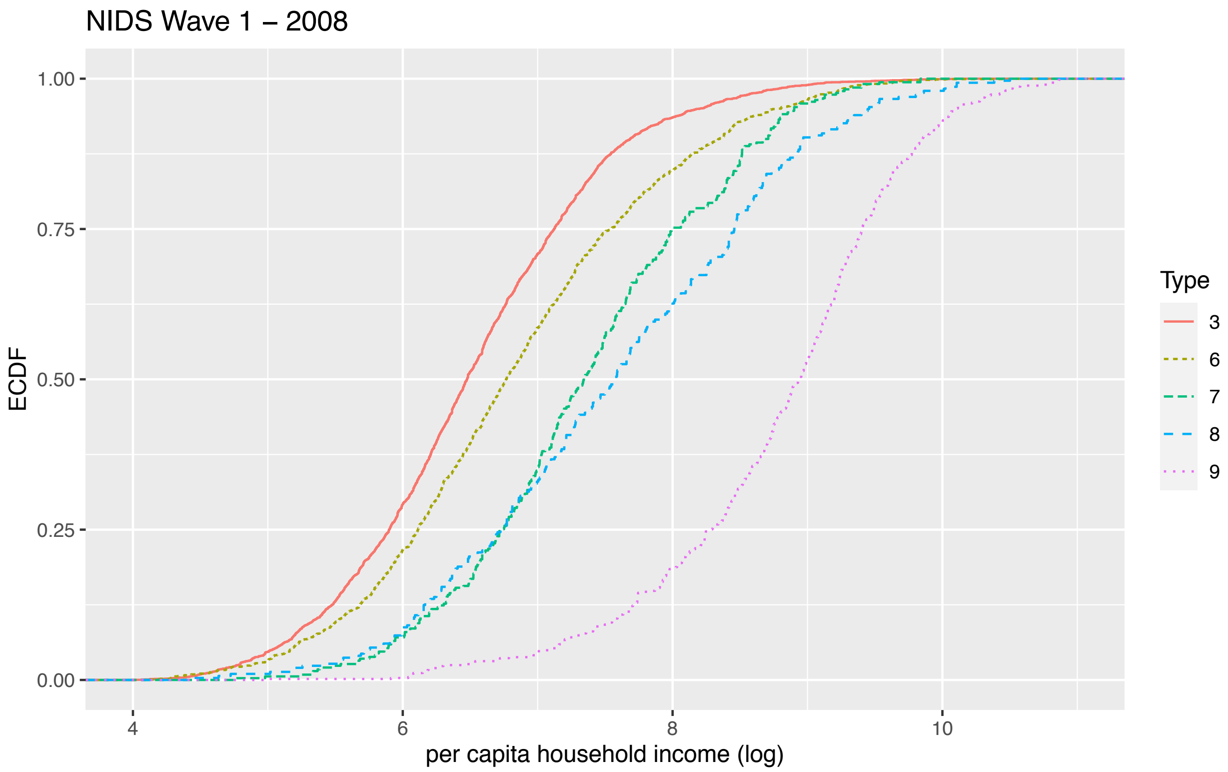
Figure 4: The partition of South Africa into five Roemerian types



Source: Own elaboration from NIDS Wave 1

Note: monthly disposable per capita incomes are in RAND 2015. Race is coded in four categories: African, Asian/Indian, Coloured, White. Parental occupation (f.occ, m.occ) is 1 digit ISCO code (10 = "never worked"). Parental education (f.edu, m.edu): 0 = no education, 1 = foundation phase, 2 = intermediate phase, 3 = senior phase, 4 = higher education.

Figure 5: Empirical cumulative distribution functions for five types in South Africa: 2008



Source: Own elaboration from NIDS Wave 1

Note: monthly disposable per capita log incomes are in RAND 2015. Types correspond to the terminal nodes of the conditional inference regression tree in Figure 4.

Given the significance test and other restrictions imposed (and described above), the regression tree arguably yields a partition of the population into the most salient types. The five selected types, corresponding to the terminal nodes in Figure 4, are: Type "3": Africans and "Coloured" people whose fathers had low levels of education (54% of the population): Type "6": Africans whose fathers had better levels of education, but were employed in lower-ranked occupations; Type "7": "Coloured" people whose fathers had better levels of education, but were employed in lower-ranked occupations; Type "8": Africans and "Coloured" people whose fathers had better levels of education and were employed in higher-ranked occupations; and Type "9": Whites and people reporting Asian/Indian ethnicity. Table 5 reports opportunity-profiles, that is population share and average outcome in each type.

Table 5: Type income means, population shares and sample sizes, over time

	2008				2010/11		
Type	**Income**	**Share**	**N**	**Type**	**Income**	**Share**	**N**
3	1,510	53.6	4,582	3	1,573	56.5	3,160
6	2,278	22.8	1,521	6	2,543	24.4	1,022
7	3,731	3.3	340	8	3,838	3.6	149
8	4,724	5.5	298	7	4,040	3.9	255
9	11,867	14.8	612	9	12,140	11.7	229
	2012				2014/15		
Type	**Income**	**Share**	**N**	**Type**	**Income**	**Share**	**N**
3	1,936	57	3,839	3	2,126	53.5	3,924
6	2,743	22.6	1,185	7	3,124	4.1	358
7	2,946	3.5	304	6	3,194	26.5	1,576
8	4,457	4.8	242	8	6,765	5.5	300
9	12,797	12.1	326	9	12,063	10.5	285
	2017						
Type	**Income**	**Share**	**N**				
3	2,164	54.5	2,752				
6	3,658	25.3	1,009				
7	4,723	3.8	243				
8	8,364	5.7	210				
9	12,218	10.6	183				

Source NIDS Wave 1 - Wave 5

Note: Types' number correspond to terminal nodes in Figure 4. Population shares are in percent, monthly disposable per capita incomes are in RAND 2015.

We define ten tranches as the tenths in the income distribution within each type (following Roemer, 1998). Based on this partition we check the dominance conditions summarized in Table 3. Each dominance is tested for each pair of waves, in a discrete number of points (10), and the result reported in one entry of a *5 x 5* matrix. If at all 10 points distribution *i* dominates distribution *j,* we report ">" in cell *i,j* (row *i*, column *j*)*.* In addition, if at all points we can reject the null hypothesis that distribution *j* dominates distribution *i* and, at least in one point we cannot reject the null hypothesis that distribution *i* dominates distribution *j,* we consistently report the confidence level*.* Table 6 reports results for simple Generalized Lorenz dominance of the overall distribution in year *i* ($w(x)$) over year *j* ($w(y)$): $GL(w(x), N) > GL(w(y), N)$ These correspond to pure prioritarian rankings, with unanimity in the class of social welfare functions S^P.

Table 6: Population-wide Generalized Lorenz dominance

	2008	**2010/11**	**2012**	**2014/15**	**2017**
2008	.	.	<***	<***	<***
2010/11	.	.	<***	<***	<***
2012	>***	>***	.	<***	<***
2014/15	>***	>***	>***	.	.
2017	>***	>***	>***	.	.

Source NIDS Wave 1 - Wave 5

*Note: confidence levels are calculated based on the percentile distribution of 500 bootstrap replications of the statistics: * = 0.9, ** = 0.95, *** =0.99.*

Table 6 indicates that social welfare in 2008 and 2010/11 cannot be unambiguously ranked in this class, as there is no second-order dominance. Neither can 2014/15 and 2017 be ranked. Aside from those two, every other pairwise comparison yields unambiguous welfare rankings: 2012 dominates both 2008 and 2010/11. 2014/15 dominates all three previous waves; and 2017 does the same, except for 2014/15. Overall, this is a tale of consistent improvements in social welfare over time (aside from a blip in 2010/11, and an inconclusive comparison between the last two waves), consistent with a rising mean and declining inequality.

As noted in Section 3, the comparisons in Table 6 are based purely on the well-being (here: income) vector and take no account of differences in circumstances and efforts. Table 7 presents results for the six different opportunity-prioritarian dominance criteria summarized in Table 3.[44]

[44] Note that Table 7 does not present the combination of ex post compensation using tranches and agnostic reward.

Table 7: Six versions of opportunity-prioritarian dominance results for South Africa: 2008-2017

a) Generalized Lorenz Dominance of the total type well-being distribution

	2008	2010/11	2012	2014/15	2017
2008	.	.	<***	<***	<***
2010/11	.	.	<***	<***	<***
2012	>***	>***	.	<***	<***
2014/15	>***	>***	>***	.	<*
2017	>***	>***	>***	>*	.

b) Sequential second-order dominance of the type-specific Generalized Lorenz curves.

	2008	2010/11	2012	2014/15	2017
2008	.	.	<***	<***	<***
2010/11	.	.	<***	<***	<***
2012	>***	>***	.	<***	<***
2014/15	>***	>***	>***	.	.
2017	>***	>***	>***	.	.

c) Sequential first-order dominance of population-weighted type-distributions

	2008	2010/11	2012	2014/15	2017
2008	.	.			<***
2010/11	.	.	<***	<***	<***
2012	.	>***	.	.	.
2014/15	.	>***	.	.	.
2017	>***	>***	.	.	.

d) Generalized Lorenz Dominance of the distribution of class means

	2008	2010/11	2012	2014/15	2017
2008	.	.	<***	<***	<***
2010/11	.	.	<***	<***	<***
2012	>***	>***	.	.	<***
2014/15	>***	>***	.	.	<***
2017	>***	>***	>***	>***	.

e) Sequential second-order dominance of the class-specific Generalized Lorenz curves.

	2008	2010/11	2012	2014/15	2017
2008	.	.	.	<***	<***
2010/11	.	.	<***	<***	<***
2012	.	>***	.	<***	<***
2014/15	>***	>***	>***	.	.
2017	>***	>***	>***	.	.

f) Sequential first-order dominance of Class distributions

	2008	2010/11	2012	2014/15	2017
2008	.	.			<***
2010/11	.	.	<***	<***	<***
2012	.	>***	.	.	.
2014/15	.	>***	.	.	.
2017	>***	>***	.	.	.

*Source: NIDS Wave 1 - Wave 5. Note: confidence levels are calculated based on the percentile distribution of 500 bootstrap replications of the statistics: * = 0.9, ** = 0.95, *** =0.99*

Table 7 contains the key empirical results of the paper. There is a one-to-one correspondence between its six panels and the six main panels of Table 3: Panels (a), (b) and (c) in Table 7 test the dominance conditions described in the first row of Table 3, for the combination of the ex-ante version of the compensation principle with the three different versions of the reward principle (utilitarian, inequality averse and agnostic). Panels (d), (e) and (f) test the dominance conditions from the second row of Table 3, for the combination of ex-post compensation with the three different versions of the reward principle.[45] As noted earlier, the type-based results in the first three panels and the class-based results in the last three would be identical if the quantile functions (or cdf's) of the five types did not ever cross in our data. Figure A.3 in the Appendix shows (for 2012) that there are type crossings in the data – for example, belonging to type 7 (Colored, with an educated father working in low occupations) is an advantage for individuals in the lowest tranches, but not so much an advantage for individuals in the highest tranches – so the ex-ante and ex-post compensation results can differ in principle. As we see below, they do differ in practice as well, although not very substantially.

How do the "opportunity-prioritarian" dominance results under these various combinations of compensation and reward principles compare among themselves, and with the pure prioritarian rankings in Table 6? Overall, there is remarkable consistency in the number and identity of rankings among the results consistent with utilitarian and inequality-averse reward, regardless of the compensation principle. These results are also generally similar to those in Table 6. The rankings in panel (b) – ex-ante compensation and inequality-averse reward – are identical to those in Table 6 (pure prioritarianism). Panel (a) – ex-ante compensation and utilitarian reward – has the same results, and adds dominance of 2017 over 2014/15, albeit only at the 10% significance level. Compared to Table 6, panel (d) – ex-post compensation and utilitarian reward – also adds dominance of 2017 over 2014/15 but loses dominance of 2014/15 over 2012. Panel (e) – ex-post compensation and inequality-averse reward – loses the dominance of 2012 over 2008.

Results are much less similar if we wish to remain agnostic about reward. In this case, the classes of social welfare functions over which we demand unanimity are much larger, so dominance is harder to obtain. In particular, first-order – rather than second-order – dominance is required: sequentially among type distributions in the ex-ante case, and sequentially among class distributions in the ex-post. It is

[45] For the combination of ex-post compensation and agnostic reward, Section 4 had also presented a dominance result using tranche distributions, rather than classes. The empirical results for that exercise are presented in Table A.2 in the Appendix.

unsurprising, then, that panels (c) and (f) display many fewer instances where rankings are unanimous for the relevant class of SWFs. Nonetheless, dominance is still found for 2017 over 2008; and for 2017, 2014/15 and 2012 over 2010/11. Interestingly, the results are identical in the ex-ante and ex-post cases.

Taken together, the empirical results suggest that Prioritarians (that is, folks with SWFs in S^P) would have a generally positive assessment of income dynamics in South Africa over those ten years. Changes in welfare between the first two waves were ambiguous. But after that, 2012 dominated both previous waves; 2014/15 dominated all three waves that preceded it; and 2017 dominated three of the four preceding waves.

More interestingly from the point of view of our analysis, modifying the tenets of prioritarianism to normatively differentiate inequalities arising from personal responsibility from those arising from exogenous circumstances makes relatively little difference – unless one is prepared to accept social-welfare functions in which greater inequality within types or classes is regarded as a social improvement. Ruling out that degree of agnosticism about the reward principle – that is, confining our attention to the utilitarian and inequality-averse versions in panels (a), (b), (d) and (e) – it turns out that, at least in this particular application, incorporating personal responsibility considerations into prioritarianism makes little difference in terms of robust rankings, regardless of which specific versions of the compensation and reward principles one happens to favor.

6. Conclusions

This paper sought to address the question of whether Prioritarianism – the view that social welfare judgements should incorporate an explicit preference for the worst-off – can be made consistent with the normative theory of Equality of Opportunity – the view that differences in well-being arising from differences in the exercise of personal responsibility may be acceptable, whereas those arising through circumstances beyond the control of the individual must be compensated. We found that the two views, in their pure forms, are inconsistent: if a person is better-off despite humble beginnings and great disadvantage, whereas someone else is worse-off despite very favorable circumstances, largely because of lack of effort or irresponsible behavior, the two normative views might recommend different policies or interventions.

However, we also found that the two views could be combined provided one accepted two main kinds of adjustments to the principles (axioms) underpinning prioritarianism. First, the notion of impartiality, or symmetry of treatment, which is universal under Prioritarianism, needs to be restricted to specific subsets of the population; for example: to those who share the same circumstances, or to those who exert similar degrees of personal effort. Second, the notion of progressive (Pigou-Dalton) transfers must also be restricted. They should no longer be required to represent improvements over the entire domain of the population but only, once again, within smaller subsets, defined either in terms of circumstances or efforts. A strengthening of the separability axiom – to an additivity one – is also needed.

Because there are differences in the specific approaches the literature has taken to the E.Op. principles of compensation and reward, we explored six main different combinations and, in each case, identified the set of axioms that would define an appropriate class of social welfare functions.[46] Furthermore, we noted – drawing on established results in the literature – the specific conditions that must hold when comparing two distributions for it to be the case that all social welfare functions in the relevant class would rank them unanimously. It turns out that all these conditions are variants of first- or second-order stochastic dominance conditions, defined over different population subgroups and/or distributions, according to the specific combination of E.Op. principles.

Finally, in an empirical application to South Africa during the 2008-2017 period, we compared the stochastic dominance results obtained empirically for the pure prioritarianism version with those for the various different versions of what we call "opportunity-prioritarianism." Despite the *a priori* risk that the differences in approaches might yield very different rankings, it turns out that, <u>in this particular application</u>, there was remarkable agreement among the various criteria, with the exception of the more demanding conditions associated with a fully agnostic attitude to the reward principle. We hope to have provided a useful toolkit of criteria for distributional comparisons for use by analysts whose social judgements incorporate two different kinds of ethical priority: a priority for the worst-off; and a priority for combating inequalities arising from circumstances for which individuals are not responsible.

[46] The six main combinations are those arising from crossing two versions of the compensation principle (ex-ante and ex-post) with three versions of the reward principle (utilitarian, inequality-averse, and agnostic). One combination – of ex-post compensation and agnostic reward – has been approached using tranches or classes, leading to the seventh cell in Table 3.

References

Adler, M. (2018). “Prioritarianism: Room for Desert?” *Utilitas*, 30 (2): 172–97.

Adler, M. (2019). *Measuring Social Welfare: An Introduction*. Oxford: Oxford University Press.

Arneson, R. (1989). “Equality of Opportunity for Welfare.” *Philosophical Studies*, 56: 77–93.

Atkinson, A. B. (1970). “On the measurement of inequality.” *Journal of Economic Theory*, 2: 244–63.

Atkinson, A. B. and Bourguignon, F. (1987). “Income Distribution and Differences in Need.” In G. F. Feiwel, ed., *Arrow and the Foundation of the Theory of Economic Policy*. London: Macmillan, pp. 350–70.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne and Son.

Brophy T., Branson, N., Daniels, R.C., Leibbrandt, M., Mlatsheni, C. and Woolard, I. (2018). *National Income Dynamics Study Panel User Manual.* Release 2018 Version 1. www.datafirst.uct.ac.za

Brunori, P., Hufe, and D. Mahler. (2018). “The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Random Forests”, Ifo Working Paper, No. 252, Ifo Institute - Leibniz Institute for Economic Research at the University of Munich, Munich.

Brunori, P., Peragine, V. and Serlenga, L. (2019). “Upward and downward bias when measuring inequality of opportunity.” *Social Choice and Welfare*, 52 (4): 635–61.

Cohen, G. A. (1989). “On the Currency of Egalitarian Justice.” *Ethics*, 99: 906–44.

Dworkin, R. (1981a). “What is equality? Part 1: Equality of welfare.” *Philosophy & Public Affairs*, 10: 185–246.

Dworkin, R. (1981b). “What is equality? Part 2: Equality of resources.” *Philosophy & Public Affairs*, 10: 283–345.

Ferreira, F. H. G. and Peragine, V. (2016). “Individual Responsibility and Equality of Opportunity”. In M. Adler and M. Fleurbaey, eds., *Oxford Handbook of Well-Being and Public Policy*. Oxford: Oxford University Press, pp. 746–84.

Fleurbaey, M. (1994). “On fair compensation.” *Theory and Decision*, 36: 277–307.

Fleurbaey, M. (1995a). “Three Solutions for the Compensation Problem.” *Journal of Economic Theory*, 65(2): 505–21.

Fleurbaey, M. and Peragine, V. (2013). “Ex ante versus ex post equality of opportunity,” *Economica,* 80: 118–30.

Fleurbaey, M., Peragine, V. and Ramos, X. (2017). “Ex-Post Inequality of Opportunity Comparisons.” *Social Choice and Welfare,* 49 (3–4): 577–603.

Hothorn T., Hornik, K. and Zeileis, A. (2006). “Unbiased Recursive Partitioning: A Conditional Inference Framework.” *Journal of Computational and Graphical Statistics*, 15(3): 651–74.

Parfit, D. (2000). “Equality or Priority?” In M. Clayton and A. Williams, eds., *The Ideal of Equality.* Houndmills: Palgrave, pp. 81–125.

Peragine, V. (2004). “Ranking income distributions according to equality of opportunity.” *Journal of Economic Inequality*, 2: 11–30.

Peragine V., Palmisano, F. and Brunori, P. (2014). “Economic Growth and Equality of Opportunity.” *World Bank Economic Review*, 28(2): 247–81.

Peragine, V. and Serlenga, L. (2008). “Higher Education and Equality of Opportunity in Italy.” In J. Bishop and B. Zheng, eds., *Research on Economic Inequality*, 16: 67–97.

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

Roemer, J. (1993). “A Pragmatic Theory of Responsibility for the Egalitarian Planner.” *Philosophy & Public Affairs*, 10: 146–66.

Roemer, J. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Saposnik, R. (1981). “Rank-dominance in Income Distributions.” *Public Choice*, 36: 147–51

Sen, A. K. (1980). “Equality of what?” In S. McMurrin, ed., *The Tanner Lectures on Human Values*. Salt Lake City: University of Utah Press, pp. 197-220.

Sen, A. K. (1985). *Commodities and Capabilities*. Amsterdam: North-Holland.

Sen, A. K. (2000). “Social Justice and Distribution of Income.” In Atkinson, A.B. and F. Bourguignon, eds., *Handbook of Income Distribution: Vol.1*. Amsterdam: North Holland, chapter 1.

Shorrocks, A. (1983). “Ranking Income Distributions.” *Economica*, 50 (197): 3–17.

Van de Gaer, D. (1993). “Equality of opportunity and investment in human capital.” Ph.D. Dissertation, Katholieke Universiteit Leuven.

Appendix A.1: Stochastic Dominance and Welfare Comparison Results

Definition of First Order Stochastic Dominance: *For any ordered distributions* ***w(x)*** *= ($w_1(x)$, …, $w_N(x)$) and* ***w(y)*** *= ($w_1(y)$, …, $w_N(y)$) with population N, distribution* ***w(x)*** *First Order Dominates distribution* ***w(y)*** *if and only if:*

$$w_k(x) \geq w_k(y),\ \ k = 1, \dots, N\ \ and\ a\ strict\ inequality\ for\ some\ k \qquad (A.1.1)$$

Equivalently, first order dominance can be expressed in terms of cumulative distribution functions: for any distributions F(x) and G(x), F first-order stochastically dominates G if and only if F (x) ≤ G (x) (for all x, and strictly for some x).

Denote by $\boldsymbol{S^I}$ the family of all social welfare functions satisfying the axioms of Strong monotonicity, Anonymity and Separability. Saposnik (1981) demonstrates the following equivalence result:

Theorem: *For any ordered distributions* ***w(x)*** *= ($w_1(x)$, …, $w_N(x)$) and* ***w(y)*** *= ($w_1(y)$, …, $w_N(y)$) with population N,*

$$S(x) > S(y) \text{ for all } S \text{ in the family } \boldsymbol{S^I} \Leftrightarrow w_k(x) \geq w_k(y),\ \ k = 1, \dots, N\ \ and\ w_k(x) > w_k(y),$$

for some *k*.

Definition of the Generalized Lorenz Curve: *For any ordered distributions X = ($w_1(x)$, …, $w_N(x)$) with population N, the Generalized Lorenz Curve of X is defined as:*

$$GL(k, x, N) = \frac{1}{N}\sum_{l=1}^{k} w_l(x),\ \ k = 1, \dots, N \qquad (A.1.2)$$

Definition of Generalized Lorenz Dominance: *For any ordered distributions* ***w(x)*** *= ($w_1(x)$, …, $w_N(x)$) and* ***w(y)*** *= ($w_1(y)$, …, $w_N(y)$) with population N, distribution* ***w(x)*** *Generalized Lorenz Dominates distribution* ***w(y)*** *($GL(w(x), N) > GL(w(y), N)$) if and only if:*

$$\frac{1}{N}\sum_{l=1}^{k} w_l(x) \geq \frac{1}{N}\sum_{l=1}^{k} w_l(y),\ for\ \ k = 1, \dots, N\ and\ a\ strict\ inequality\ for\ some\ k \qquad (A.1.3)$$

Denote by $\boldsymbol{S^S}$ the family of all social welfare functions satisfying the axioms of Strong monotonicity, Anonymity, Separability and Strong Pigou Dalton Transfer. A corollary of Theorem 2 in Shorrocks (1983) states that all *S* in $\boldsymbol{S^S}$ will rank ***w(x)*** as preferable to ***w(y)*** if and only if the distribution ***w(x)*** displays generalized Lorenz dominance over ***w(y):***

Shorrocks Theorem: *For any ordered distributions* ***w(x)*** *= ($w_1(x)$, …, $w_N(x)$) and* ***w(y)*** *= ($w_1(y)$, …, $w_N(y)$) with population N,*

$$S(x) > S(y) \text{ for all } S \text{ in the family } \boldsymbol{S^S} \Leftrightarrow GL(w(x), N) > GL(w(y), N) \qquad (A.1.4)$$

Appendix A.2: Summary of results in Fleurbaey, Peragine and Ramos (2017)

Formally, given a distribution *x*, let us denote the class *i* by $C_i(x)$, $i = 1, \dots, n$, with corresponding mean $\mu(C_i(x))$, population $n(C_i(x))$, cumulative distribution function $F_i(w_{ij}(x)|C_i)$ and Generalized Lorenz Curve $GL(C_i, x)$. Fleurbaey et al. (2017) obtain the following set of results:

i. FPR (2017) first criterion, based on utilitarian reward, implies a sequential procedure of comparison of the class distributions, where, at each step, the distributions are simply ranked by their means. This condition can also be interpreted as Generalized Lorenz dominance of the distributions of the Class means. Formally, $S(x) > S(y)$ if and only if

$$\sum_{i=1}^{k} \mu\big(C_i(x)\big) > \sum_{i=1}^{k} \mu\big(C_i(y)\big), \quad for\ \ k = 1, \dots, n \qquad (A.2.1)$$

ii. FPR (2017) second criterion, based on agnostic reward, implies a sequential procedure of comparison of the class distributions, where, at each step, the distributions are ranked by first order dominance. Formally, $S(x) > S(y)$ if and only if

$$\sum_{i=1}^{k} \big[F_i\big(w_{ij}(x)|C_i\big)\big] > \sum_{i=1}^{k} \big[F_i\big(w_{ij}(y)|C_i\big)\big], \qquad for\ k = 1, \dots, n \qquad (A.2.2)$$

iii. FPR (2017) third criterion, based on inequality averse reward, implies a sequential procedure of comparison of the class distributions, where, at each step, the distributions are ranked by Generalized Lorenz dominance. Formally, $S(x) > S(y)$ if and only if

$$\sum_{i=1}^{k} [GL(\,x, C_i)] > \sum_{i=1}^{k} [GL(y, C_i)], \qquad for\ \ k = 1, \dots, n \qquad (A.2.3)$$

Appendix A.3: Empirical analysis: Descriptive statistics and illustrations.

As discussed in Section 5, the National Income Dynamics Study (NIDS) Panel Survey contains information on various individual characteristics that can be confidently treated as circumstances, in an E.Op. sense. Unfortunately, however, information on these variables is frequently missing, as documented in Table A.1.

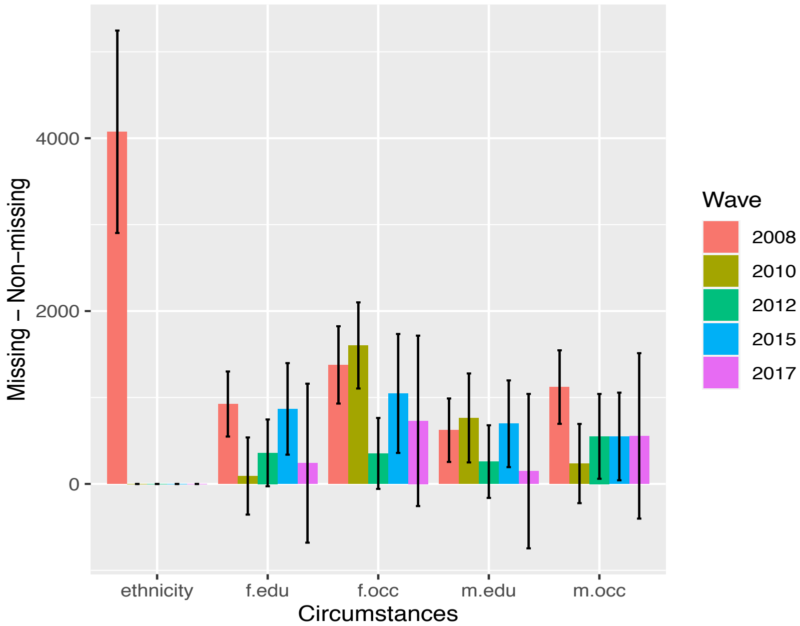
Table A.1: Distribution of missing information

Wave	Father occupation (f.occ)	Mother occupation (m.occ)	Father education (f.edu)	Mother education (m.edu)	Ethnicity
2008	22%	24%	31%	33%	4%
2010	29%	29%	42%	36%	0%
2012	23%	25%	30%	32%	0%
2014	27%	29%	29%	33%	0%
2017	36%	38%	41%	45%	0%

Source NIDS Wave 1 - Wave 5

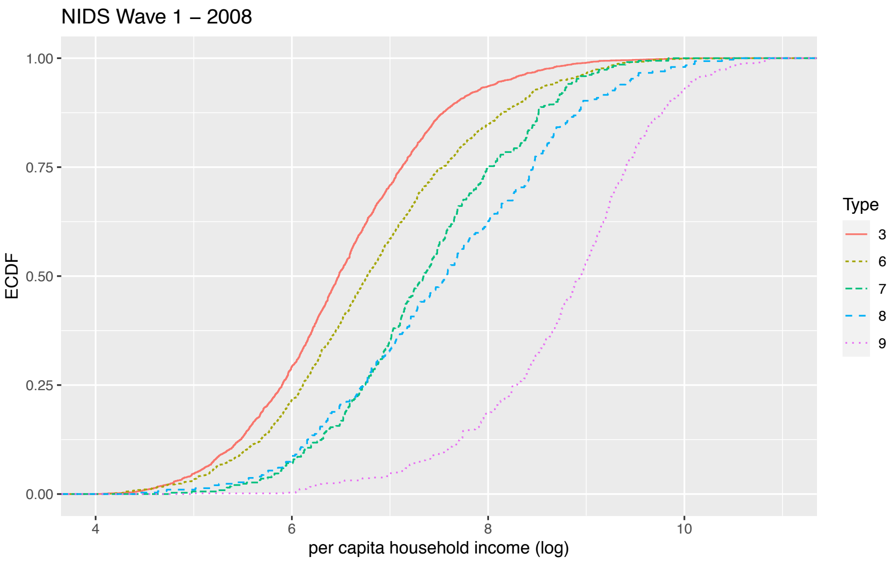
As might be expected, missing observations are not randomly distributed: there is selection in item non-response. In particular, there is a positive correlation between the outcome variable of interest, namely household per capita income, and the probability that an observation has missing information on a circumstance. Figure A. 1 plots the difference between average household disposable income for observations with and without missing information circumstances. The difference is statistically significant for all circumstances in 2008 and never statistically significant in 2017. The particularly alarming correlation detected for race in 2008 involves a relatively small number of observations (4%).

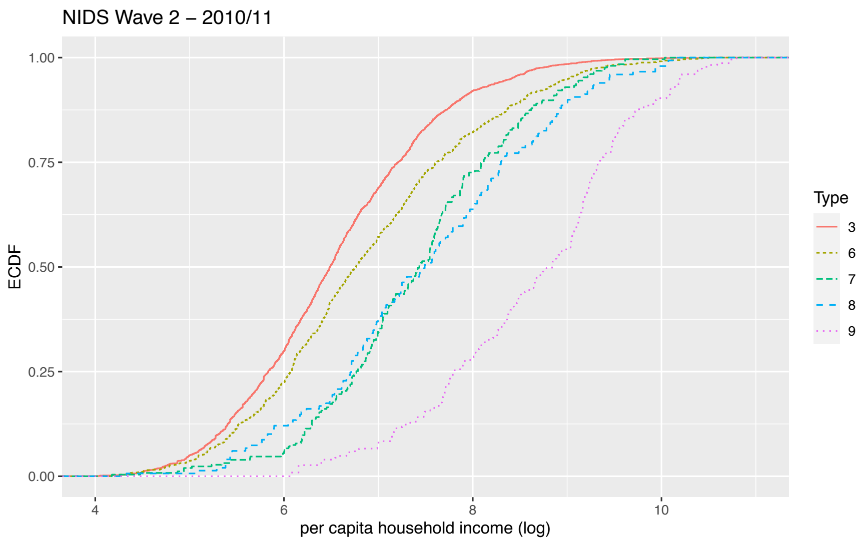
Figure A.1: Income differences between observations with missing and non-missing circumstances

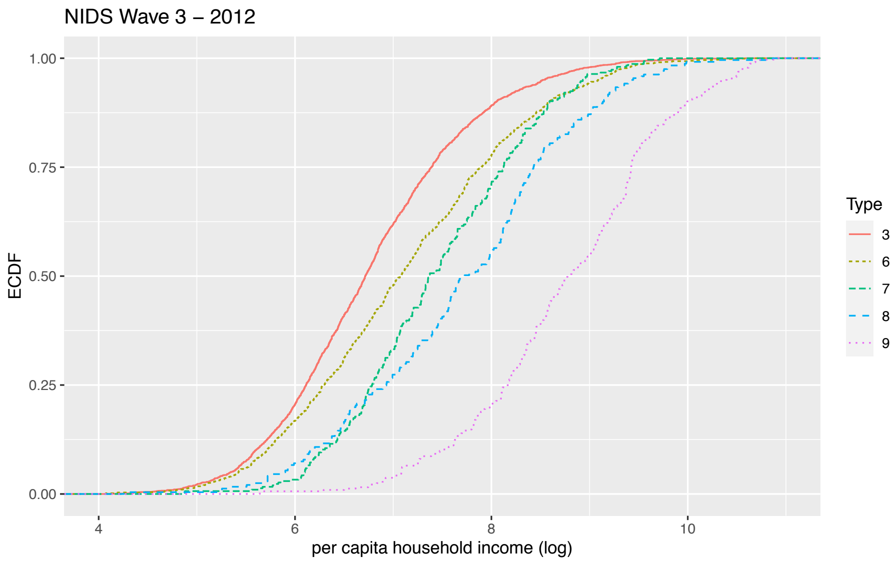


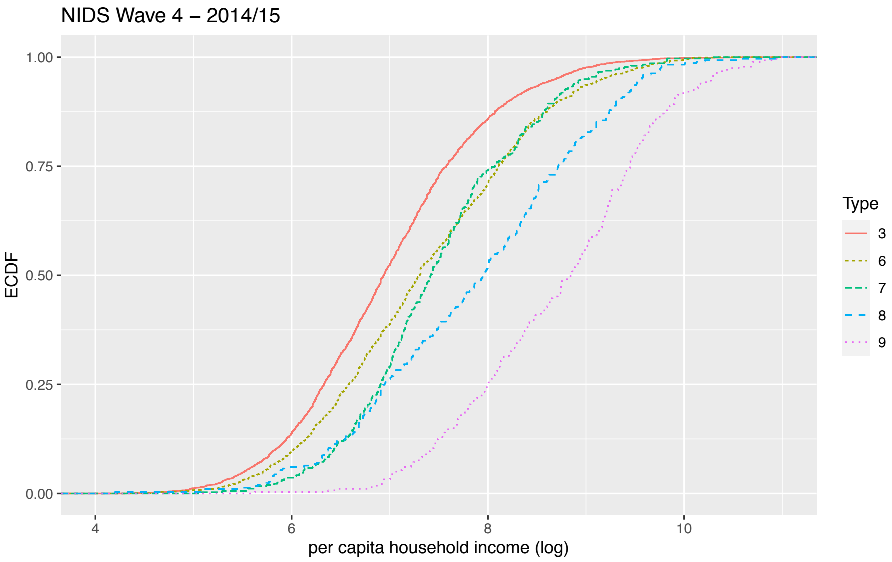
Source NIDS W1-W5

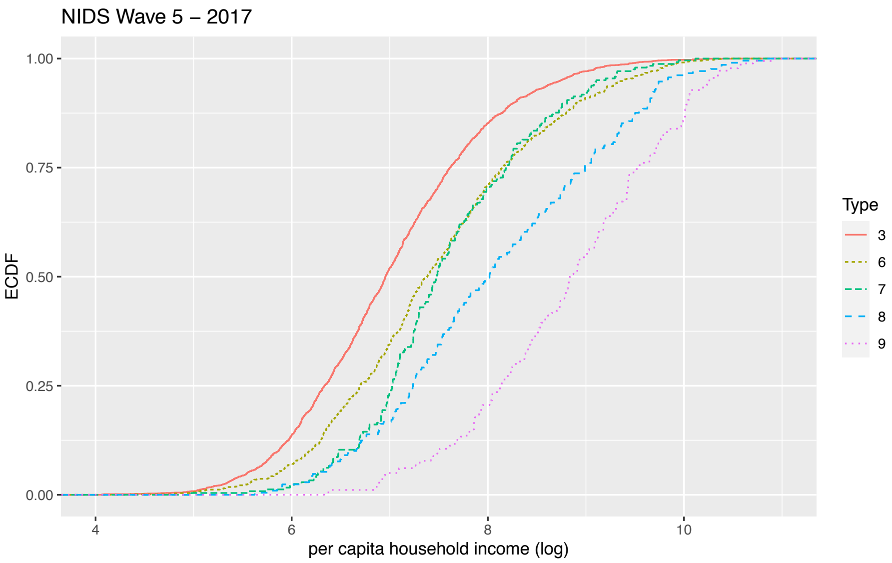
Figure A.2 – A.6: Empirical cumulative distribution functions for five types in South Africa: 2008-2017




NIDS Wave 2 – 2010/11
ECDF
1.00
0.75
0.50
0.25
0.00
4
6
8
10
per capita household income (log)
Type
3
6
7
8
9

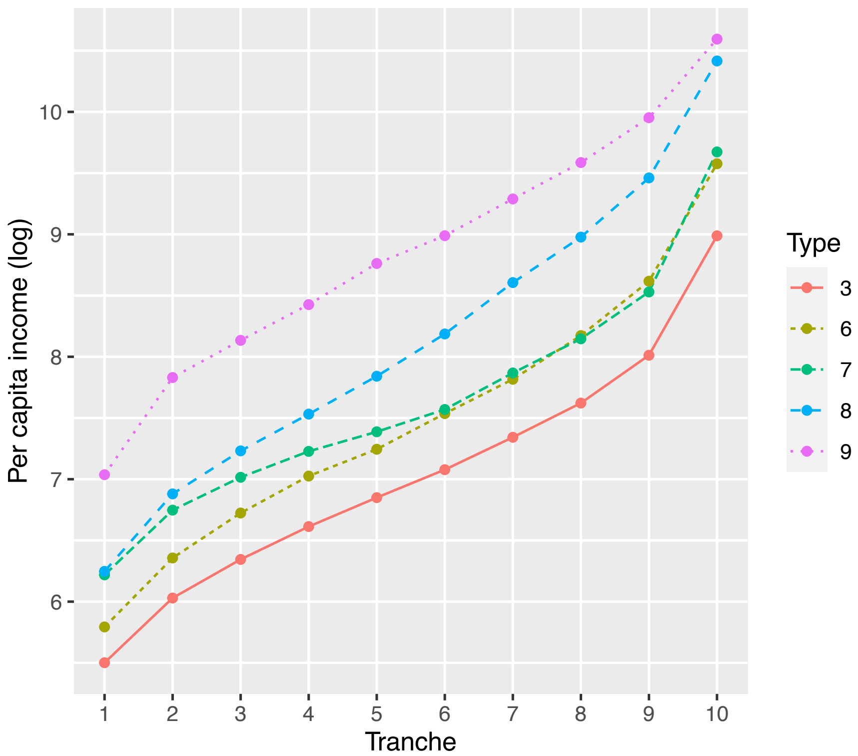

NIDS Wave 3 – 2012
ECDF
1.00
0.75
0.50
0.25
0.00
4
6
8
10
per capita household income (log)
Type
3
6
7
8
9





Source NIDS W1-W5

Figure A.7: Re-ranking of types across tranches in 2012



Source: NIDS Wave 3

Table A.2: Tranches generalized Lorenz dominance

	2008	2010/11	2012	2014/15	2017
2008	.	.	.	.	<***
2010/11	.	.	<***	<***	<***
2012	.	>***	.	<***	<***
2014/15	.	>***	>***	.	.
2017	>***	>***	>***	.	.

Source NIDS Wave 1 - Wave 5

*Note: confidence levels are calculated based on the percentile distribution of 500 bootstrap replications of the statistics: * = 0.9, ** = 0.95, *** =0.99.*