



Working Paper Series

**Measuring multidimensional inequality
of opportunity**

Martyna Kobus
Marek Kapera
Vito Peragine

ECINEQ WP 2020 - 528

Measuring multidimensional inequality of opportunity

Martyna Kobus

Polish Academy of Sciences

Marek Kapera

Polish Academy of Sciences

Vito Peragine

University of Bari

Abstract

This paper develops a normative approach to the measurement of ex-ante inequality of opportunity in a multidimensional setting, i.e., when the individual outcome is represented by a multidimensional variable. We characterize three classes of social welfare functions, all endorsing ex ante compensation but each of them reflecting a specific reward principle: (1) utilitarian, (2) agnostic and (3) averse. The first class is implemented via generalized Lorenz Dominance applied to each attribute separately. The agnostic and inequality averse classes are implemented by a welfarist Lorenz ordering, namely, of type-aggregate utilities. In the case of inequality-averse class, utility functions are submodular, hence capturing the dependence between attributes. We also develop normative inequality indices (Atkinson, 1970; Kolm 1969; Sen, 1973) for the classes of welfare functions and study their properties. Finally, we propose an empirical applications of the methods developed in the paper: by using the National Longitudinal Study of Adolescent to Adult Health (Add Health) we evaluate inequality of opportunity in U.S. for the case of three dimensions of individual outcomes, namely, education, health and income.

Keywords: Equality of opportunity; economic inequality; multidimensional welfare; multidimensional inequality; Lorenz dominance.

JEL Classification: D63; I32.

1 Introduction

Equality of opportunity (EOp or, conversely, IOp for inequality of opportunity) is a widely held ideal of fairness, stating that differences in life success should reflect differences in individual effort, but should not be determined by factors beyond the control of the individual, the so-called circumstances. Building on a distinguished literature in political philosophy (e.g. Arneson 1989; Cohen 1989; Dworkin 1981; Rawls 1971; Roemer 1993), the concept has recently stimulated a growing body of research in economics and related disciplines (see Fleurbaey 2008; Roemer 1998). These ideas have also influenced agenda-setting and policy-making at the highest levels. Goal 10 of the United Nations Sustainable Development Goals (SDGs), for example, recognizes that ‘income inequality cannot be effectively tackled unless the underlying inequality of opportunities is addressed’. Alan Krueger, a former main economic adviser to President Obama, once said ‘The rise in inequality in the United States over the last three decades has reached the point that inequality in incomes is causing an unhealthy division in opportunities, and is a threat to our economic growth.’ Equality of opportunity, as opposed to pure equality of outcomes, offers a version of egalitarianism that recognizes the role of individual responsibility.

The ideal of equality of opportunity rests on two basic principles (Fleurbaey 2008): the compensation principle, stating that inequalities arising from individual circumstances (factors beyond the individual responsibility) are unfair and should be compensated by society; and the reward principle, which addresses the problem of how to apportion outcomes to different degrees of effort. Different versions of these two principles have been proposed in the literature. In particular, ex-post compensation (Roemer 1993) states that inequalities across people with different circumstances should be equalized after effort is realized. This is mainly achieved by equating outcomes of those who exert the same level of effort. On the other hand, ex-ante compensation (Van de Gaer 1993) states that inequalities across people with different circumstances should be equalized before effort is realized. This is achieved by equating opportunity sets.¹ This latter goal has been operationalized by interpreting the outcome distribution among individuals with the same circumstances (so-called ‘types’ in Roemer’s terminology) as the opportunity sets open to each individual in that group. Hence the focus is on the (reduction of) inequality between types. As far as the reward principle is concerned, different versions of the principle have been proposed by the literature, expressing different attitudes with respect to the outcome inequality observed among individuals endowed with the same circumstances: from utilitarian reward (Fleurbaey 2008; Van de Gaer 1993), which expresses perfect neutrality, to inequality-averse reward (Ramos and Van de Gaer 2016), which expresses aversion to inequality, to intermediate and agnostic positions (Fleurbaey and Peragine 2013).

The theoretical and the empirical literature on EOp has grown enormously in recent years, and a vast array of methods and approaches to the measurement of IOp has been explored: see Ferreira and Peragine (2015), Ramos and Van de Gaer (2016), and Roemer and Trannoy (2015) for recent surveys. However, in most of the existing literature, both theoretical and empirical, the outcome of interest is typically represented by a unidimensional variable: income, consumption, education, or health. On the other hand, both the theory and the practice of inequality measurement have moved towards the multidimensional space. For a long time, many researchers have advocated the inclusion of non-income dimensions in the evaluation of well-being (Atkinson and Bourguignon 1982; Kolm 1977; Maasoumi 1986; Sen 1973). Apparently, their call was heeded: for instance, the United Nations now uses a multidimensional measure, the Human Development Index, to evaluate the well-being of nations and their progress over time. Several multidimensional measures of welfare and of inequality have been developed (Gajdos and Weymark 2005; Maasoumi 1986; Tsui 1995, 1999); these are measures based on joint distributions that are sensitive to multidimensional generalizations of the Pigou–Dalton transfer (PDT) principles,

¹ Fleurbaey and Peragine (2013) explore the potential conflicts between the ex-ante and ex-post approaches to compensation and between different versions of compensation and the reward principle.

such as uniform PDT or uniform majorization (Tsui 1999), and to transfers that change the dependence structure—namely, so-called correlation-increasing switches (Tsui 1999). Thus, it is widely acknowledged in the literature and more so in practical decision-making that well-being is a multidimensional concept and cannot be reduced to a single proxy such as income.

Therefore, if the goal of public policy is to seek EOp, its measurement has to account for the multidimensionality of well-being. To date, however, researchers on equality of opportunity have made only a little progress on this mission.

Some empirical works jointly investigate the presence of unequal opportunities for multiple outcomes, by treating those outcomes as separate identities. Examples include: Bourguignon et al. (2007), who analyse income and schooling outcomes in Brazil; Ferreira and Gignoux (2011), who focus on different income measures; and Peragine and Serlenga (2008), who analyse university graduation results and later-life income. In addition to these works, Ferreira and Gignoux (2010) estimate IOp in Turkey by considering PISA (Programme for International Student Assessment) scores in reading, math, and science as outcomes of interest. The reports by Paes de Barros et al. (2009) and Son (2013) go beyond the focus on income and educational outcomes by assessing equal opportunities in income, education, and access to basic services such as sanitation for a sample of Latin American and Asian countries, respectively. In spite of differences with respect to the outcome dimensions, the populations of interest and the methodological approaches towards IOp measurement, these works treat the issue of multidimensionality in the same simplistic manner: each dimension of interest is analysed separately, while neglecting any interdependencies between the outcomes of interest. Some other works attempt to circumvent multidimensionality by focusing on a unique indicator that arguably captures the relevant information for the outcome of interest (see Yalonetzky 2012). For instance, it is acknowledged that health is a multidimensional outcome that lends itself to empirical analysis from an equal-opportunity perspective. For the purpose of reducing the dimensionality of the analytical problem, most works focus on self-assessed health as the appropriate *equilizandum* (Li Donni et al. 2011; Rosa Dias 2009; Trannoy et al. 2010). In spite of the well-acknowledged merits of this measure (Idler and Kasl 1995), the focus on a single measure entails an information loss that could be countervailed by developing an appropriate multidimensional measure of EOp.

To summarize, to date the existing EOp literature has not yet combined ‘the two aspects, individual responsibility (translated into seeking equality in the domain of opportunities) and multidimensionality of well-being’ (Ferreira and Peragine 2015) in a satisfying manner. The present paper aims at filling this gap: we develop a normative approach to the measurement of IOp in a multidimensional setting—that is, when the outcome of interest is a multidimensional variable.

We focus on the *ex-ante* approach. Hence, in our model within each type there is a multidimensional outcome distribution. Formally, types differ not only dimension by dimension, but also with respect to the dependence structure between dimensions. In other words, in a multidimensional setting the effect of circumstances may not only be that worse types have worse distributions of outcomes, but also that individuals in worse types are more likely to be deprived *jointly* in several outcomes than in better types. The latter concerns dependence, which is a truly distinctive feature of multidimensionality. What is the extent of unfair opportunities if one takes into account both marginal distributions of well-being dimensions and their dependence? Or, how does one measure IOp when joint distribution of outcomes is considered? We develop tools that allow us to answer such questions. To lay the foundation of such measurement, we use an axiomatic approach.

We consider three classes of welfare functions that endorse the *ex-ante* compensation principle and cover major reward principles (i.e. neutral, agnostic, and averse) and link these classes to implementable criteria (i.e. different versions of Lorenz dominance). Then we construct IOp measures that are induced

by these three classes via the so-called AKS transformation (Atkinson 1970; Kolm 1969; Sen 1973). A more detailed description of the setting and the results follows.

We have an ordering of the society (multidimensional) outcome distributions, which is represented by a continuous social evaluation function. We start by characterizing the class of such functions on the basis of axioms that express different versions of the reward and the ex-ante compensation principle. These axioms use basic transformations that define our concepts of inequality. In a multidimensional setting, the defining inequality transformations are transformations that reduce spread in many dimensions (i.e. PDT on each dimension with potentially different amounts of attributes transferred) and transformations that change the dependence between dimensions (i.e. correlation-increasing or -decreasing switches—see, for example, Tsui 1999). Different reward principles reflect different sensitivity to such transformations when they happen *within a type*; in particular, a welfare function which respects utilitarian reward does not change, since it does not pay attention to any inequality within type; a welfare function which is agnostic is silent; and the one which respects averse reward goes up.

In Theorem 1 we show that the preference relation represented by the class of evaluation functions that are monotone, additive (i.e. sum utilities within types and between types), symmetric, neutral to inequality within type (i.e. obey the utilitarian reward principle) but averse to inequality between types (i.e. obey the ex-ante compensation principle) is implemented via CGLD, which applies standard generalized Lorenz dominance to each attribute separately and to a distribution such that each individual in their type possesses the same amount of each attribute (i.e. perfect within-type equality). We further extend this result to allow for more general aggregation procedures than additivity (Theorem 2). Then, in Theorems 3 and 4 we characterize classes of evaluation functions that are, respectively, agnostic and averse to inequality *within* type. Here, the implementable criterion is Lorenz dominance applied to the sum of individual within-types utilities, where utility functions are described by a set of properties. This definition of Lorenz dominance is welfaristic in a sense that it applies to the case in which individual well-being is first evaluated by a utility function, similarly to the approach taken by, for example, Maa-soumi (1986) and Dardanoni (1992). The better-off type has higher average welfare—that is, welfare is understood as the sum of individual utilities within a type. For the case of aversion, Lorenz dominance acts on a class of utility functions that are increasing, type-concave, concave, and submodular. The two latter properties reflect the multidimensionality aspect. In the case of submodular functions it is assumed that dimensions of well-being are substitutes; an extension to the case of complementary goods is straightforward.

For inequality measures, the class generated by the class of welfare functions that is neutral to within-type inequality is of a very simple form, namely the weighted sum of types' means in each attribute normalized by the value of the highest welfare. The latter is obtained for a distribution which is perfectly equal within and between types—that is, a matrix of population means. The weights assigned to types preserve aversion to inequality between types—that is, worse type gets higher weight. When the class of welfare functions used is averse to inequality within type, one cannot give an explicit formula for related class of inequality measures; however, with the addition of a well-known property of ratio scale invariance (Tsui 1995), we obtain a specific class of measures. Ratio scale invariance means that attributes can be rescaled in any way; for example, it does not matter whether income is measured in dollars or hundreds of dollars. The measures obtained with the previous class are only weakly relative—that is, the attributes can only be scaled by the same factor. The within-type inequality-averse class of measures consists of two components: one that reflects the aversion to inequality in the distribution of attributes within a type (as measured by Tsui's (1995) equality index); and another which reflects aversion to inequality between types' aggregate utilities. This means that when we put the value of Tsui's (1995) equality index to 1, we obtain a multidimensional IOp measure that focuses only on between-type inequality. The parameters of this class reflect the dimension weights. This class of measures takes into account the whole multidimensional distribution within a type, and in particular it is sensitive to how

dependent the dimensions of well-being are. This result is particularly interesting from the point of view of empirical applications.

Finally, we relate the obtained IOp measures to Lorenz dominance conditions. In the case of neutrality within type, inequality indices are implemented by CGLD applied to outcome matrices normalized by the value of highest welfare—that is, for the perfectly equal distribution. Recall that in the standard case (Atkinson 1970) inequality and welfare rankings are intertwined, providing that a certain condition, such as equality of means, is preserved. In other words, welfare ranking between distributions is the same as inequality ranking between distributions normalized by the mean. Here, the welfare ranking of the distributions is the same as the inequality ranking of the distributions normalized by the highest welfare. For the case of inequality aversion within types, inequality ranking is implemented by a welfaristic Lorenz dominance that applies to the same class of utility functions as in Theorem 4 for inequality-averse welfare functions, and applies to outcome matrices normalized by the vector of attributes' means.

We close the paper with an empirical application based on the National Longitudinal Study of Adolescent to Adult Health (Add Health) in the USA, waves 1–4. We use self-reported personal earnings, years of education, and body mass index (BMI) as outcomes and a set of circumstance variables. Aversion to within-type inequality matters significantly in the evaluation of IOp. It constitutes more than half of overall IOp. The results show how considering the dimensions jointly as opposed to considering them separately significantly changes the evaluation of IOp. For example, univariate IOp for education and BMI are, respectively, 0.06 and 0.021, whereas for income it is 0.47. In such a case, IOp when outcomes are considered jointly equals 0.2. This comes from the fact that joint evaluation takes into account that (1) some outcomes are distributed more equally than others; (2) rankings of types with respect to mean outcomes values differ across outcomes (i.e. best income type is not necessarily best BMI type); and (3) within type, individuals occupy different positions on respective dimensions (i.e. highly educated individuals are not necessarily individuals with the highest incomes). These are all factors that are missing when the focus is limited to a single well-being outcome such as income.

The paper is structured as follows. Section 2 introduces the analytical framework. In Section 3 the basic definitions and axioms are introduced and discussed. Section 4 contains the characterization theorems for welfare functions, linking them to Lorenz dominance conditions, with each class as a separate subsection. The same structure continues in Section 5, where we state the results for classes of inequality measures. Finally (Section 6), we link inequality measures to dominance conditions similar to those introduced in Section 4. Section 7 applies the developed methods to the US data. There are many extensions of the current framework and we discuss them all in Section 8, which concludes the paper.

2 The analytical framework

We follow Peragine (2004) and extend his ex-ante approach by allowing a multidimensional distribution of outcomes. We have a society consisting of N individuals, where each individual is completely described by an outcome x , which in the present paper will be treated as a multidimensional variable consisting of k attributes, and a set of traits that can be divided into: traits beyond the responsibility of individuals, called circumstances and represented by set variable $O \in \mathcal{O}$; and those for which the individual is fully responsible, called effort and represented by $w \in \Theta \subseteq \mathbb{R}_+$. We assume that $\mathcal{O} = \{O_1, O_2, \dots, O_n\}$ and its elements are sorted according to ordering \prec , that is $O_i \prec O_{i+1}$ for $i \in \{1, 2, \dots, n-1\}$. Furthermore, we assume that all individuals have the same degree of access to the set Θ of possible values of effort. Outcome is generated by a function $g : \mathcal{O} \times \Theta \rightarrow \mathbb{R}_+^k$ that assigns individual outcomes to

combinations of circumstances and effort: $x = g(O, w)$.² We assume that it is monotone in both circumstances and effort, that is $g(O_i, w) < g(O_{i+1}, w)$ for $i \in 1, 2, \dots, n-1$, $w \in \Theta$, $g(O_i, w_1) < g(O_i, w_2)$ for $i \in 1, 2, \dots, n$, and $w_1 < w_2 \in \Theta$. By $<$ we mean inequality with respect to each coordinate in \mathbb{R}_+^k .

Given the ordering $<$, we can divide the society according to individuals' circumstances, namely into n types. We say that people in the same type have equal opportunities, as determined by the set of circumstances they are given. Let N_h be the number of people in type h . We have $\sum_{h=1}^n N_h = N$. We denote by $X^h \in M_{N_h \times k}$ the type h profile, that is outcome distribution within type h . X_i^h, X_j^h are, respectively, rows and columns of matrix X^h , and X_{ij}^h represents an outcome for a j th attribute for an i th individual who sits in type h . Matrices X^h are arranged into matrix X in increasing order, that is rows $1, \dots, N_1$ of X are rows of X^1 , rows $N_1 + 1, \dots, N_1 + N_2$ are rows of X^2 , and so on. Finally, let X_μ be a matrix of type means. Keep in mind that it represents the case of perfect within-type equality, but *not* between-type equality. The distribution that represents both perfect within- and between-type equality is denoted X^μ —that is, it is a matrix of population means.³ We have $X^\mu, X_\mu \in M_{N \times k}$ —that is, these are matrices such that we still maintain the same number of rows for each type.

Let $D = \{X \in M_{N \times k}(\mathbb{R}_+) : g \text{ is monotone in circumstances and effort}\}$ denote the set of possible outcome profiles. We are interested in the ranking of outcome matrices from the set D —that is, we examine a binary preference relation of an opportunity egalitarian social decision-maker. We assume that this relation is a continuous ordering, hence it has its representation via a continuous social evaluation function $W : D \rightarrow \mathbb{R}$. $I : D \rightarrow [0, 1]$ is a corresponding inequality index. Formally, the relationship between W and I is defined in the following way.

Definition 1. We say that an inequality measure I_W is induced by a welfare function W if $I_W(X) = 1 - \delta(X)$ where $\delta(X) \in [0, 1]$ satisfies equation $W(X) = W(\delta(X)X^\mu)$.

The properties of both W and I will be delineated in the next sections. This is the so-called Atkinson–Kolm–Sen approach (Tsui 1995) to constructing inequality measures, which ensures that they are normatively significant—that is, under certain restrictions (e.g. equality of means or total sums) $W(X) \leq W(Y) \iff I(X) \geq I(Y)$.

3 Definitions and axioms

We start by defining a general multidimensional transfer. This definition will then be used in subsequent definitions and axioms imposed on the function W .

Definition 2. Transfer between types (TT). Let $X, Y \in D$ and let $O_s < O_r$. We say that $X = A_\varepsilon(Y, r, s, p, q)$ is obtained by transfer between types (TT) from Y if $X^h = Y^h$ for all $h \notin \{r, s\}$ and $X_i^l = Y_i^l$ for $l \in \{r, s\}$ and $i \notin \{p, q\}$, with $X_{pj}^r = Y_{pj}^r \varepsilon_j + Y_{qj}^s (1 - \varepsilon_j)$ and $X_{qj}^s = Y_{qj}^s (1 - \varepsilon_j) + Y_{pj}^r \varepsilon_j$ where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)$ with $\varepsilon_j \geq 0$.

Here $A_\varepsilon(Y, r, s, p, q)$ is a succinct way to contain information about the most general transformation, namely, (i) a transfer is made between types r, s , (ii) potentially different amounts of each attribute are transferred, i.e., ε is a vector with potentially different elements, and (iii) ε can be greater than 1,

² In other words, given a combination of effort and circumstances, function g completely determines the outcomes—that is, all individuals with a given effort–circumstance bundle share the same outcomes. Richer models of EOP also include a ‘luck component’ (Lefranc et al. 2008).

³ Although we do not deal with effort directly, in this framework if we needed we could also adopt Roemer’s (1998) statistical solution. This solutions says that people who are at the same percentile of their outcome distribution exercised comparable degree of effort, only here it is a percentile of a multidimensional distribution.

therefore Definition 2 defines both a progressive and a regressive transfer from r to s . That is, depending on the magnitudes of Y^r, Y^s and the value of ε sometimes the attributes are transferred from r to s and sometimes they are transferred to r from s . If we want to ensure that a progressive transfer is made when $Y_{pj}^r > Y_{qj}^s$, and a regressive transfer is made when $Y_{pj}^r < Y_{qj}^s$, then we need to control the sign of $1 - \varepsilon_j$. This is summarized in the following definition.

Definition 3. Pigou–Dalton transfer between types (PDTT). Let $O_s \prec O_r$. For $X, Y \in D$, we say that X is obtained from Y by means of Pigou–Dalton transfer between types (PDTT) if X is obtained by TT from Y and $\text{sgn}(1 - \varepsilon_j) = \text{sgn}(Y_{pj}^r - Y_{qj}^s)$ holds—that is, ε is a transfer from type r to s .

Please note that in Definition 3, when the value of the attribute in the better type is greater than the value of the attribute in the worse type $Y_{pj}^r > Y_{qj}^s$, $\text{sgn}(Y_{pj}^r - Y_{qj}^s) > 0$ and thus $\varepsilon < 1$, we transfer from r to s and this is a progressive transfer i.e. a transfer from a higher amount of attribute to a lower amount of attribute. Similarly, when the value of the attribute in the better type is lower than the value of the attribute in the worse type $Y_{pj}^r < Y_{qj}^s$, $\text{sgn}(Y_{pj}^r - Y_{qj}^s) < 0$ and thus $\varepsilon > 1$, we transfer from r to s and this is a regressive transfer i.e. a transfer from a lower amount of attribute to a higher amount of attribute.

When $r = s$ the transfer takes place within the same type, but again allows for potentially *different proportions* of each attribute to be transferred. A restriction on ε makes this a transfer from the richer to the poorer individual within a given type. This is expressed in Definition 4, which is our key definition for spread-reducing transfers.

Definition 4. Pigou–Dalton transfer (PDT). For $X, Y \in D$, we say that X is obtained from Y by means of PDT if $X = A_\varepsilon(Y, l, l, p, q)$, where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)$ with at least one $\varepsilon_i > 0$.

In a multidimensional setting, welfare and inequality are not only about the spread, but also about the dependence structure between attributes. This is reflected in the definition of a correlation-increasing transfer (Boland and Proschan 1988), where between two individuals one is assigned maximum values of each attribute and the other is assigned minimum values. Such an operation clearly increases correlation between dimensions. Although marginal distributions of attributes do not change, there is now ‘more joint risk’—that is, there is higher likelihood that a given individual occupies higher or lower positions in several dimensions. This is formally defined in Definition 5.

Definition 5. Correlation-increasing transfer (CIT). For all $X, Y \in D$, if $X^h = Y^h$ for all $h \neq l$, we say that Y^l is obtained from X^l by the CIT if $Y^l = C(X^l, p, m)$, (with $p < m$) where $C(X^l, p, m) \in M_{N_l^X \times k}$, $c_{pj} = \max(y_{pj}, y_{mj})$, $c_{mj} = \min(y_{pj}, y_{mj})$ and $c_{ij} = y_{ij}$ elsewhere.

These definitions will now be the basis of the axioms that are imposed on the function W . These axioms are as follows.

Monotonicity (MON). For all $X, Y \in D$, if $X^h = Y^h$ for all $h \neq m$, $X_{pq}^m = Y_{pq}^m + \varepsilon$ with $\varepsilon > 0$ and $X_{ij}^l = Y_{ij}^l$ for $(i, j) \neq (p, q)$ then $W(X) \geq W(Y)$.

Additivity (ADD). There exist functions $U^h : \mathbb{R}^k \rightarrow \mathbb{R}$, for all $h \in \{1, 2, \dots, n\}$, assumed to be twice differentiable (almost everywhere) in the variable X_i^h , such that $W(X) = \sum_{h=1}^n \sum_{i=1}^{N_h^X} U^h(X_i^h)$ for all $X \in D$.

Inequality neutrality within types (INWT). For all $X, Y \in D$, if X is obtained from Y by PDT, then $W(X) = W(Y)$.

For all $X, Y \in D$, if $X^h = Y^h$ for all $h \neq l$, $X^l = C(Y^l, p, m)$, where $C(\cdot, p, m)$ is CIT for some $p < m$ then $W(X) = W(Y)$.

Inequality agnosticism within types (IAGWT). For all $X, Y \in D$, $X^h = Y^h$ for all $h \neq l$, X^l is obtained from Y^l by PDT, then W is salient.

For all $X, Y \in D$, $X_i = Y_i$ for all $i \neq l \in \{1, 2, \dots, n\}$, $X^l = C(Y^l, h, m)$, where $C(\cdot, h, m)$ is CIT for some $h < m$, then W is salient.

Inequality aversion within types (IAWT). For all $X, Y \in D$, if $X^s = Y^s$ for all $i \neq l \in \{1, 2, \dots, n\}$, Y^l is obtained from X^l by PDT, then $W(X) < W(Y)$.

For all $X, Y \in D$, if $X_i = Y_i$ for all $i \neq l \in \{1, 2, \dots, n\}$, $X^l = C(Y^l, h, m)$, where $C(\cdot, h, m)$ is CIT for some $h < m$ then $W(X) < W(Y)$.

Inequality aversion between types (IABT) Let $X, Y \in D$. Further, let $O_s \prec O_r$, that is $r \neq s$. Then, the monotonicity of function g ensures that for each effort level ω , r has better outcomes than s . Now let X be a PDTT of Y . Then $W(X) > W(Y)$.

MON and ADD are technical properties and widely assumed in the literature. In ADD we define U^h as independent of individuals in type h ; this would be implied if we assumed axiomatically type symmetry, namely, that function W is invariant with respect to transformations of individual bundles *within* a type, but not between types (as in standard symmetry), because for the EOp framework it does matter from which type an individual is. MON states that welfare function increases following an increase in the value of one attribute for a single individual. ADD means that we consider utilitarian welfare functions—that is, W is a sum of utilities. Please note that an individual utility function is a function of many variables here. INWT, IAGWT, IAWT, and IABT are conceptual and specific to the EOp framework. The first three are reward principles. INWT means that W is invariant to inequality-reducing transformations within type, that is PDT and CIT. INWT expresses the utilitarian reward principle (see Van de Gaer 1993), and hence expresses neutrality with respect to inequality within types. On the other hand, IAGWT assumes nothing (i.e. it is agnostic) and IAWT expresses aversion with respect to inequality within types (inequality-averse reward principle). IABT is an ex-ante principle of compensation (e.g. Fleurbaey and Peragine 2013; Peragine 2004), which says that a transfer from a better to a worse type increases welfare.

We relax the assumption of additivity at one stage of aggregation of utilities; namely, we allow for general aggregation between types.

Separability between types (SBT). There exists function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$, and for all $h = 1, \dots, n$ there exist functions $u_h : \mathbb{R}^{N_h^X} \rightarrow \mathbb{R}$ and $U^h : \mathbb{R}^k \rightarrow \mathbb{R}$ assumed to be twice differentiable (almost everywhere), such that $u_h = \sum_{i=1}^{N_h^X} U^h(X_i^h)$ and $W(X) = \psi(u_1, \dots, u_n)$.

The ethical requirement imposed on the aggregation procedure ψ is the one that expresses the principle of compensation: welfare does not decrease following a (cardinal) welfare transfer from the more opportunity-endowed to the less opportunity-endowed types.

Inequality aversion between types' welfare (IABT*). Let $X, Y \in D$. Further, let $O_s \prec O_r$, that is $r \neq s$ and $\varepsilon > 0$. We have

$$\psi(u_1, \dots, u_n) < \psi(u_1, \dots, u_s + \varepsilon, \dots, u_r - \varepsilon, \dots, u_n).$$

4 Welfare functions and dominance conditions

In this section we study three classes of social welfare functions and for each of them we give a dominance condition which implements the ordering prescribed by a given class.

4.1 Neutrality

We start with the following class:

$$\mathcal{W}^{AOEN} = \{W | \text{MON, ADD, INWT, IABT}\}.$$

AOEN means ‘additive opportunity egalitarian neutral’, where the last term comes from neutrality embedded in INWT.

INWT brings us asymptotically to a matrix that has mean value for each attribute in a type; that is, type distribution can be summarized by a vector of means. Indeed, in the case of INWT a social decision-maker does not care about the distribution of outcomes within a type. This is expressed formally in Lemma 1.

Lemma 1. *Let $X \in D$ and W satisfies INWT. Then $W(X) = W(X_\mu)$, where $(X_\mu)_{ij}^h = \frac{\sum_{i=1}^{N_h} X_{ij}^h}{N_h}$.*

Proof. By INWT we know that $W(X) = W(Y)$ where Y is PDT of X . We will prove that X_μ is the result of a sequence of PDT transformations of X . It suffices to show it for a given type h . We start with averaging the values of attributes for individual $i = 1$ by the following sequence of transfers. Let $(X_1)^h = X^h$ and $(X_{l+1})^h = A_\varepsilon((X_l)^h, h, h, 1, l + 1)$, where $\varepsilon_{l+1} = (\frac{l}{l+1}, \dots, \frac{l}{l+1})$. That is, in each step we make a PDT transfer between individual $i = 1$ and individual l . We obtain that $X_{1j}^h = \sum_{l=1}^{N_h} \frac{X_{l,j}^h}{N_h}$ and then we apply the same procedure to the next individual. \square

We now define component-wise Lorenz dominance. It performs standard Lorenz dominance for each attribute separately—that is, it works on vector columns of matrix X .

Definition 6. **Component-wise Lorenz dominance (CLD).** *For all increasingly ordered (within columns) matrices $X, Y \in M_{N \times k}(\mathbb{R}_+)$, we say that X component-wise Lorenz dominates Y ; that is*

$$X \succeq_{CLD} Y \iff \frac{\sum_{i=1}^l X_{ij}}{\sum_{i=1}^N X_{ij}} \geq \frac{\sum_{i=1}^l Y_{ij}}{\sum_{i=1}^N Y_{ij}} \quad \forall 1 \leq l \leq N \forall 1 \leq j \leq k$$

CLD can be generalized to the case when columns in both matrices do not have the same total sum.

Definition 7. **Component-wise generalized Lorenz dominance (CGLD).** *For all increasingly ordered (within columns) matrices $X, Y \in M_{N \times k}(\mathbb{R}_+)$, we say that X component-wise generalized Lorenz dominates Y ; that is*

$$X \succeq_{CGLD} Y \iff \sum_{i=1}^l X_{ij} \geq \sum_{i=1}^l Y_{ij} \quad \forall 1 \leq l \leq N \forall 1 \leq j \leq k$$

We can now relate CGLD to a class of functions \mathcal{W}^{AOEN} .

Theorem 1. *For $X, Y \in D$ we have*

$$X_\mu \succeq_{CGLD} Y_\mu \iff W(X) \geq W(Y) \quad \text{for all } W \in \mathcal{W}^{AOEN}$$

Proof. By ADD we have $W(X) = \sum_{h=1}^n \sum_{i=1}^{N_h} U^h(X_i^h)$. We now show that $U^h(X_i^h) = \sum_{j=1}^k U_j^h(X_{ij}^h)$, $dU^h/dX > 0$, and D^2U^h is a zero matrix. MON implies $dU^h/dX > 0$. Further, let $X_{i_1}^h, X_{i_2}^h$ be two outcomes. From INWT and ADD, we have that $U^h(X_{i_1}^h) + U^h(X_{i_2}^h) = U^h(X_{i_1}^h + \varepsilon \cdot (X_{i_2}^h - X_{i_1}^h)) + U^h(X_{i_2}^h - \varepsilon \cdot (X_{i_2}^h - X_{i_1}^h))$ (by \cdot we denote coordinate-wise multiplication of vectors). It is equivalent to $\frac{U^h(X_{i_1}^h) - U^h(X_{i_1}^h + \varepsilon \cdot (X_{i_2}^h - X_{i_1}^h))}{|\varepsilon \cdot (X_{i_2}^h - X_{i_1}^h)|} = \frac{U^h(X_{i_2}^h - \varepsilon) - U^h(X_{i_2}^h)}{|\varepsilon \cdot (X_{i_2}^h - X_{i_1}^h)|}$. Moving to zero in the limit with the norm of ε , we obtain

equality of derivatives for arbitrary points $X_{i_1}^h, X_{i_2}^h$, and therefore $U^h(X_i^h) = \sum_{j=1}^k U_j^h(X_{i_j}^h)$, D^2U^h as a zero matrix. In the reverse implication, this functional form of W clearly satisfies ADD and MON. Since every operation was invertible, we obtain that equality of gradients of U^h is equivalent to INWT.

Now to show that $dU^h/dX > dU^{h+1}/dX \geq 0$, we use IABT. Let us assume that $X_p^h < X_q^{h+1}$. Since function is additive, without loss of generality, we consider W which is a function of first dimension only. Then by IABT we have $U_1^h(X_{p1}^h + \varepsilon \cdot (Y_i^h - X_i^h)) - U_1^h(X_{p1}^h) > U_1^{h+1}(X_1^{h+1}) - U_1^{h+1}(X_1^{h+1} + \varepsilon \cdot (Y_i^h - X_i^h))$. Dividing it by $|\varepsilon \cdot (Y_i^h - X_i^h)|$ and taking it to 0 in the limit gives us the result.

Now let $X_\mu \succeq_{CGLD} Y_\mu$. Using Theorem 1 (Peragine 2004) for any j and the first part of this proof, we get that $W(X) \geq W(Y)$. Now by contradiction, let us assume that $W(X) \geq W(Y)$ and $X_\mu \not\succeq_{CGLD} Y_\mu$. Then let us take $W(X) = W_j(X)$, where W_j is W restricted to j th dimension, which belongs to \mathcal{W}^{AOEN} . This contradicts Theorem 1 (Peragine 2004). \square

Corollary 1. Let $W \in \mathcal{W}^{AOEN}$. Then W is of the form

$$W(X) = Nb + \sum_{h=1}^n \sum_{i=1}^{N_h} \sum_{j=1}^k a_j^h X_{ij}^h$$

with $a_j^h > a_j^{h+1}$ for all j .

Proof. From proof of Theorem 1 we know that $W(X) = \sum_{h=1}^n \sum_{i=1}^{N_h} \sum_{j=1}^k U_j^h(X_{ij}^h)$, and that gradients of U_j^h are constant. From this, we obtain that each U_j^h is a linear function. Condition on coefficients a_j^h follows from $dU^h/dX > dU^{h+1}/dX$. \square

Theorem 1 states that if for each attribute the type-means distribution of X generalized Lorenz dominates the type-means distribution of Y , then for all welfare functions in the class \mathcal{W}^{AOEN} X is more opportunity egalitarian than Y . The converse is also true. Please note that the definition of Lorenz dominance applies to increasingly ordered vectors, which means that in each column means (i.e. columns of X_μ) are increasingly ordered. The interpretation of Lorenz dominance is meaningful only if the ordering is the same in each column, otherwise it could be the case that the type which is considered better has lower mean than a worse type for some attributes. In what follows we remove the constants. Corollary 1 further specifies the form of welfare functions that are in the class characterized by Theorem 1. These functions are a weighted sum of utility functions that are linear and have higher weights for worse types (to preserve IABT). Theorem 1 gives a simple prescription of how to implement unanimous welfare rankings for the case of neutrality. First partition the society into types, then for each attribute compute the distribution of means. Finally, apply generalized Lorenz dominance to the obtained distribution. The result is restrictive—each dimension is treated separately, and therefore the obtained ranking has a meaningful interpretation only when a better type has a higher mean than a worse type on each dimension. This restrictiveness is driven by the powerful axiom INWT.

We will now extend Theorem 1 by allowing more general welfare aggregation procedures. That is, we keep additivity for a within-type aggregation of individual utilities; however, we relax the assumption of additivity of types. To this end, we define class $\mathcal{W}^{OEN} = \{W | \text{MON, INWT, IABT*}, \text{SBT}\}$. Provided that one accepts the IABT* axiom, unambiguous social ranking can be achieved in the case of general aggregation of the types welfare.

Theorem 2. For $X, Y \in D$ we have $W(X) \geq W(Y)$ for any $W \in \mathcal{W}^{OEN}$ if and only if $X_\mu \succeq_{CGLD} Y_\mu$.

Proof. Since $\mathcal{W}^{AOEN} \subset \mathcal{W}^{OEN}$, then by Theorem 1 $W(X) \geq W(Y)$ for any $W \in \mathcal{W}^{OEN}$ implies $X_\mu \succeq_{CGLD} Y_\mu$, so we only have to prove the second implication.

Let us assume $X_\mu \succeq_{CGLD} Y_\mu$. By defining $X^* = (u_1, \dots, u_N)$, $Y^* = (v_1, \dots, v_n)$ with $u_i = U^h(X_i^h)$, $v_i = U^h(Y_i^h)$ we obtain by Theorem 2 (Peragine 2004) that $W(X) \geq W(Y)$ if and only if $X_\mu^* \succeq_{CGLD} Y_\mu^*$. Hence we obtain that $X_\mu \succeq_{CGLD} Y_\mu$ implies $W(X) \geq W(Y)$ for any $W \in \mathcal{W}^{OEN}$ if and only if $X_\mu \succeq_{CGLD} Y_\mu$ implies $X_\mu^* \succeq_{CGLD} Y_\mu^*$.

We have $\sum_{i=1}^h (X_\mu)_{ij} \geq \sum_{i=1}^h (Y_\mu)_{ij}, \forall 1 \leq h \leq n, \forall 1 \leq j \leq k$. Assuming that U_j^h and ϕ are increasing, we get

$$\sum_{i=1}^l u_i \geq \sum_{i=1}^l v_i \quad \forall 1 \leq l \leq n$$

which is equivalent to $X_\mu^* \succeq_{CGLD} Y_\mu^*$. From MON we obtain that W is increasing in every variable. Since ϕ is increasing, it means that U_j^h are also. □

4.2 Agnosticism

The CGLD characterized in the previous section is non-welfarist in the sense that the evaluation of distributions depends *directly* on the values of the attributes. Different authors (e.g. Dardanoni 1992; Maasoumi 1986) propose a welfarist approach to multidimensional aggregation. Namely, an individual well-being derived from attributes is first evaluated through a utility function. This leads to the utility-based majorization, such as the one below (see e.g. Tsui 1999).

Definition 8. Lorenz dominance For all $h = 1, \dots, n$ let $u_h^X = \sum_{i=1}^{N_h^X} U^h(X_i^h)$ (as in Definition 3). For all increasingly ordered vectors, X strictly Lorenz dominates Y in class of functions \mathcal{F} , that is

$$X \succeq_{LD(\mathcal{F})} Y \iff \sum_{h=1}^l u_h^X \geq \sum_{h=1}^l u_h^Y \quad \forall l=1, \dots, n, \forall U^h \in \mathcal{F}$$

Lorenz dominance first aggregates individual utilities within type, and then compares partial sums of such within-type aggregate utility vectors. Here we make the dependence on \mathcal{F} explicit to underlie that Lorenz dominance will be applied to different classes of utility functions.

We define the class of welfare functions that fulfil the agnostic version of the reward principle. Here, AOEAG means ‘additive opportunity egalitarian agnostic’. We have

$$\mathcal{W}^{AOEAG} = \{W | \text{MON, ADD, IABT}\}$$

and the following class of utility functions

$$\mathcal{U}^{ICLA} = \{U | \text{Increasing, Type-Concave}\}$$

where by increasingness we understand $dU^h/dX > 0$ —that is, the function is increasing with respect to attributes and type-concavity is defined as follows.

Definition 9. Type-concavity. Function $U^h : \mathbb{R}^k \rightarrow \mathbb{R}$ is type-concave if the better the type the lower its first derivative. Formally, $dU^h/dX > dU^{h+1}/dX > 0$.

Theorem 3 combines the class of welfare functions \mathcal{W}^{AOEAG} and Lorenz dominance applied to the class \mathcal{U}^{ICLA} .

Theorem 3. $X \succeq_{LD(\mathcal{U}^{ICLA})} Y \iff W(X) \geq W(Y)$ for all $W \in \mathcal{W}^{AOEAG}$

Proof. We start by showing that $W \in \mathcal{W}^{AOEAG}$ if and only if it is of form

$$W(X) = \sum_{h=1}^n \sum_{i=1}^{N_h^X} U^h(X_i^h)$$

with $dU^h/dX > dU^{h+1}/dX > 0$. Obviously, ADD is equivalent to $W(X) = \sum_{h=1}^n \sum_{i=1}^{N_h^X} U^h(X_i^h)$ and MON is equivalent to $dU_i^h/dX > 0$. We only need to show that IABT is equivalent to $dU^h/dX > dU^{h+1}/dX$.

Let us assume that $r < s \in \{1, \dots, n\}$, $p \in \{1, \dots, N_r^X\}$, $q \in \{1, \dots, N_s^X\}$. We set $\varepsilon = (\varepsilon_1, 1, 1, \dots, 1)$. Then by IABT we have $\frac{U^r(X_p^r \cdot \varepsilon + X_q^s \cdot (1-\varepsilon)) - U^r(X_p^r)}{1-\varepsilon_1} > \frac{U^s(X_q^s) - U^s(X_q^s \cdot \varepsilon + X_p^r \cdot (1-\varepsilon))}{1-\varepsilon_1}$ where by \cdot we describe coordinate-wise multiplication of vectors. Treating $X_{\cdot 1}$ as variable (first coordinate of vector X), we now obtain $dU^r/dX_{\cdot 1} > dU^s/dX_{\cdot 1}$ by moving ε_1 to 1 in the limit. By setting $\varepsilon = (1, \dots, 1, \varepsilon_l, 1, \dots, 1)$ we obtain the same result for arbitrary coordinates, which concludes this part of the proof.

Now, let us assume that $X \succeq_{LD(\mathcal{Q}^{ICAL})} Y$. Then $\sum_{h=1}^l u_h^X \geq \sum_{h=1}^l u_h^Y \quad \forall l=1, \dots, n$. Since it applies for any $l \in \{1, \dots, n\}$ then especially for $l = n$, so

$$\sum_{h=1}^n u_h^X \geq \sum_{h=1}^n u_h^Y \iff \sum_{h=1}^n \sum_{i=1}^{N_h^X} U^h(X_i^h) \geq \sum_{h=1}^n \sum_{i=1}^{N_h^Y} U^h(Y_i^h) \iff W(X) \geq W(Y)$$

We now turn to the second implication. We have $W(X) \geq W(Y)$ for any $W \in \mathcal{W}^{AEOAG}$, and we want to prove that $X \succeq_{LD(\mathcal{Q}^{ICAL})} Y$. In other words, we have $\sum_{h=1}^n u_h^X \geq \sum_{h=1}^n u_h^Y$ and want to prove that this implies $\sum_{h=1}^l u_h^X \geq \sum_{h=1}^l u_h^Y$ for any $l \in \{1, \dots, n\}$.

We proceed by contradiction. Let us assume that for some $l \in \{1, \dots, n\}$, $\sum_{h=1}^l u_h^X < \sum_{h=1}^l u_h^Y$. For $l = n$, $W(X) < W(Y)$, which is a contradiction. For $l < n$, let us denote $A = \sum_{h=1}^l u_h^Y - \sum_{h=1}^l u_h^X > 0$. Define W' with $U'^h = U^h$ for $h \leq l$ and $U'^h = \frac{A}{2(\sum_{h=l+1}^n u_h^Y - \sum_{h=l+1}^n u_h^X)} U^h$. Since $\sum_{h=l+1}^n u_h^Y - \sum_{h=l+1}^n u_h^X > 0$ and $A > 0$ such W' is in \mathcal{W}^{AEOAG} and $\sum_{h=1}^n u_h^X - \sum_{h=1}^n u_h^Y = \frac{A}{2} - A < 0$, which contradicts assumption that $\sum_{h=1}^n u_h^X \geq \sum_{h=1}^n u_h^Y$. \square

Theorem 3 states that in order to compare distributions of outcomes in terms of welfare functions that are monotone, additive, and agnostic with respect to inequality within type but averse to inequality between types, one can apply Lorenz dominance to type-aggregate utilities, where utility functions are increasing and type-concave. Indeed, such utility functions are ‘salient’ about inequality within type, but respect aversion to inequality between types. The reverse is true as well, namely, Lorenz Dominance for such utility functions is the largest (in the sense of inclusion) ordering on D consistent with the class \mathcal{W}^{AEOAG} .

4.3 Aversion

Now we will turn to the case when a policy-maker cares about both, namely, inequality between types (IABT) and inequality within type (IAWT).

We define the following class of welfare functions, where AOEAE means ‘additive opportunity egalitarian averse’:

$$\mathcal{W}^{AOEAE} = \{W | \text{MON, ADD, IAWT, IABT}\}$$

and the following class of utility functions:

$$\mathcal{U}^{ICL} = \{U | \text{Increasing, Type-Concave, Component-wise Concave, Submodular}\}$$

Definition 10. Submodularity. Function U^h is submodular, if $U^h(X_p^h) + U^h(X_q^h) > U^h(X_p^h \wedge X_q^h) + U^h(X_p^h \vee X_q^h)$ where $X_p^h \wedge X_q^h$ is a vector of elements $\max\{X_{pj}^h, X_{qj}^h\}$ and $X_p^h \vee X_q^h$ of $\min\{X_{pj}^h, X_{qj}^h\}$.

Note that $X_p^h \vee X_q^h$ denotes component-wise maximum and $X_p^h \wedge X_q^h$ denotes component-wise minimum. The function is submodular if it attains lower value if, between two individuals, one has lower (or higher) value than the other for *each* attribute, than for any other distribution of attributes. Submodularity reflects that association between dimensions matters, and if there is more of it the utility is lower.

Theorem 4 combines welfare functions and dominance criterion. Here, Lorenz dominance is applied to utility aggregates where utility functions belong to class \mathcal{W}^{ICL} . It is then an implementable criterion for the class of ‘additive opportunity egalitarian averse’ welfare functions.

Theorem 4.

$$X \succeq_{LD(\mathcal{W}^{ICL})} Y \iff W(X) \geq W(Y) \text{ for all } W \in \mathcal{W}^{AOEA}$$

Proof. We start by showing that $W \in \mathcal{W}^{AOEA}$ if and only if it is of the form

$$W(X) = \sum_{h=1}^n \sum_{i=1}^{N_h^X} U^h(X_i^h)$$

with U^h being component-wise concave, type-concave, and submodular.

From Theorem 3 we must only show that IAWT is equivalent to submodularity and component-wise concavity of U^h . We start by proving that the second part of IAWT (concerning CIT) is equivalent to submodularity of U^h . Let $X, Y \in D$, $X^i = Y^i$ and $Y^l = C(X, i, i')$ for some $i \neq l \in \{1, \dots, n\}$. From IAWT we have that $W(X) > W(Y)$, from which by ADD we obtain

$$U^h(X_i^h \wedge X_{i'}^h) + U^h(X_i^h \vee X_{i'}^h) < U^h(X_i^h) + U^h(X_{i'}^h)$$

which is the definition of submodularity.

Now we need to prove that the first part of IAWT is equivalent to concavity of U^h in each dimension. First, we prove that IAWT implies concavity. Without loss of generality let $X_p^l < X_q^l$ and set ε as in definition of PDTT. Now let $X, Y \in D$, $Y_i = X_i$ for $i \neq l \in \{1, \dots, n\}$ and $Y^l = BX^l$. By IAWT we have $W(X) < W(Y)$, which is equivalent to

$$U^h(\varepsilon X_p^l + (1 - \varepsilon)X_q^l) + U^h(\varepsilon X_q^l + (1 - \varepsilon)X_p^l) > U^h(X_p^l) + U^h(X_q^l),$$

$$U^h(\varepsilon X_p^l + (1 - \varepsilon)X_q^l) - U^h(X_p^l) > U^h(X_q^l) - U^h(\varepsilon X_q^l + (1 - \varepsilon)X_p^l)$$

from

$$\varepsilon X_p^l + (1 - \varepsilon)X_q^l - X_p^l = (1 - \varepsilon)(X_q^l - X_p^l) = X_q^l - (\varepsilon X_q^l + (1 - \varepsilon)X_p^l)$$

$$\frac{U^h(\varepsilon X_p^l + (1 - \varepsilon)X_q^l) - U^h(X_p^l)}{\|(1 - \varepsilon)(X_q^l - X_p^l)\|} > \frac{U^h(X_q^l) - U^h(\varepsilon X_q^l + (1 - \varepsilon)X_p^l)}{\|(1 - \varepsilon)(X_q^l - X_p^l)\|}$$

now $\varepsilon \rightarrow 1$

$$DU^h(X_p^l) \geq DU^h(X_q^l)$$

For proof of the second implication, let us note that any PDTT can be decomposed into a sequence of one-dimensional PDTTs. Hence, we can without loss of generality assume that $\varepsilon_i = 1$ for $i \neq 1$. We denote Y as PDT of X . From IAWT then we have $W(Y) > W(X)$ and we have

$$W(Y) > W(X) \iff U^l(\varepsilon \cdot X_p^l + (1 - \varepsilon) \cdot X_q^l) + U^l(\varepsilon \cdot X_q^l + (1 - \varepsilon) \cdot X_p^l) > U^l(X_p^l) + U^l(X_q^l)$$

which follows from Jensen’s inequality. This concludes this part of the proof.

Now assume that $X \succeq_{LD(\mathcal{U}^{ICL})} Y$. In exactly the same way as in the proof of Theorem 3 we obtain that $\sum_{h=1}^n u_h^X \geq \sum_{h=1}^n u_h^Y$, so we have $W(X) \geq W(Y)$. Obviously, since $\mathcal{W}^{AEOA} \subset \mathcal{W}^{AEOAG}$, then by Theorem 3 the second implication follows. \square

The following corollary states what is clear from the proof of Theorem 4, namely that if goods are complements and not substitutes, the theorem still holds but for a different class of utility functions.

Corollary 2. In the definition of IAWT we change the sign of function W —that is, for all $X, Y \in D$ if X is obtained via CIT from Y , then $W(X) > W(Y)$ and in Definition 10 we change the sign of inequality—that is, $U^h(X_p^h) + U^h(X_q^h) > U^h(X_p^h \wedge X_q^h) + U^h(X_p^h \vee X_q^h)$ (i.e. function U^h is supermodular). Then Theorem 4 holds for

$$\mathcal{U}^{ICL} = \{U \mid \text{Increasing, Type-Concave, Component-wise Concave, Supermodular}\}.$$

5 Inequality of opportunity measures derived from welfare functions

In this section we first define the inequality of opportunity measures and then characterize, for each class of social welfare functions the corresponding classes of measures.

Definition 11. I is an IOp measure if it satisfies the following properties

1. I is continuous.
2. $I(X^\mu) = 0$, that is I is zero for perfect equality.
3. $I(Y) < I(X)$ if Y is PDTT of X .

Measure I is relative if additionally $I(XC) = I(X)$ for diagonal matrix C . If $I(XC) = I(X)$ holds only for C being diagonal matrix with equal elements on the diagonal, then we say that I is weakly relative.

In what follows we derive IOp measures for classes (1) \mathcal{W}^{AOEN} (i.e. neutrality); (2) \mathcal{W}^{OEN} (i.e. neutrality with general aggregation of types); and (3) \mathcal{W}^{AOEA} (i.e. aversion). Agnosticism puts too little structure on the functions W to derive IOp measures or to study their properties.

5.1 Neutrality

Before we characterize the IOp measures for the class of \mathcal{W}^{AOEN} welfare functions, please note that the second point of Definition 11 does not state that I is zero only for X^μ . In fact, for welfare functions satisfying INWT, the corresponding inequality measure I is zero for any X which has total sums equal everywhere to that of X^μ . That is, if the index is indifferent to inequality within type, the total sum of each attribute can be distributed in any way within type and the index should not change.

It is not clear that inequality measures induced from welfare functions $W \in \mathcal{W}^{AOEN}$ according to Definition 1 satisfy Definition 11, but with slight restriction, namely $W(X) \neq 0$, this is indeed the case. Also, since it was noted that I could be zero not only for X^μ , it is not obvious that the inequality measure induced by a welfare function is well defined; that is, that δ in Definition 1 is uniquely determined. It is true, however, as it follows from linearity (Corollary 1); that is, if $W(Y) = W(X^\mu)$, then $W(\delta Y) = \sum_{h=1}^n \sum_{i=1}^{N_h^Y} \sum_{j=1}^k a_j^h \delta Y = \delta \sum_{h=1}^n \sum_{i=1}^{N_h^Y} \sum_{j=1}^k a_j^h Y = \delta W(Y) = \delta W(X^\mu) = W(\delta X^\mu)$.

Theorem 5. Let $W \in \mathcal{W}^{AOEN}$ with $W(X) \neq 0$ for any $X \in D$. Then we have the following set of results:

1. I_W is given by $1 - \frac{W(X)}{W(X^\mu)} = 1 - \frac{\sum_h \sum_i \sum_j a_j^h X_{ij}^h}{\sum_{h=1}^n N_h \sum_{j=1}^k a_j^h (X^\mu)_{1j}^h}$.
2. I_W is given by $1 - \frac{W(X^\mu)}{W(X^\mu)} = 1 - \frac{\sum_{h=1}^n N_h \sum_{j=1}^k a_j^h (X^\mu)_{1j}^h}{\sum_{h=1}^n N_h \sum_{j=1}^k a_j^h (X^\mu)_{1j}^h}$.
3. I_W is a weakly relative inequality measure.

Proof. From Definition 1, we have $I_W(X) = 1 - \delta(X)$, where $\delta(X)$ satisfies $W(\delta(X)X^\mu) = W(X)$. Since by Theorem 1 any $W \in \mathcal{W}^{AOEN}$ is affine, we have $W(\delta(X)X^\mu) = \delta(X)W(X^\mu) = W(X) \iff \delta(X) = \frac{W(X)}{W(X^\mu)}$, proving item 1. Now by Lemma 1 we obtain item 2.

To show item 3, we check whether I is in fact a weakly relative inequality measure. Since W is the sum of affine functions, it is continuous and since $W(X^\mu) \neq 0$ so is $1 - \frac{W(X)}{W(X^\mu)}$. For fully egalitarian X of course $X^\mu = X$ so $\frac{W(X)}{W(X^\mu)} = 1$, from which $I_W(X) = 0$. By IABT we have $W(X) < W(Y)$ if Y is PDTT of X . Since PDTT does not change total sums in any dimension, we get that $X^\mu = Y^\mu$. This ends the proof that I_W is an inequality measure. Using the fact that W is affine, we obtain that I_W is weakly relative, since in this case $W(XC) = cW(X)$, where c is the value from the diagonal of C . To show that it is not a relative index, we give a counterexample. Let $n = 2, N_1^X = N_2^X = 1, j = 2$. Set $X_{11}^1 = X_{12}^1 = 1, X_{11}^2 = 1, X_{12}^2 = 3$. Also set $a_1^1 = a_2^1 = 2, a_1^2 = a_2^2 = 1$. Now $W(X) = 8, W(X^\mu) = 9$, hence $\delta(X) = \frac{8}{9}$. Set C diagonal with $c_{11} = 2, c_{22} = 1$ and we have $W(XC) = 11, W(X^\mu C) = 12$ hence $\delta(XC) = \frac{11}{12} \neq \frac{8}{9}$. \square

Theorem 5 states that the index related to the class of welfare function \mathcal{W}^{AOEN} is one minus the weighted sum of type-means for each dimension normalized by the highest amount of welfare achievable (2 in the theorem). Here types' weights a_j^h follow Corollary 1—that is, on each dimension j , we have $a_j^h > a_j^{h+1}$. Higher weights are assigned to types that have higher mean on each dimension. This measure is a *weakly* relative measure; that is, it does not change when all attributes are scaled by the same factor, but it is *not* invariant when each attribute is scaled by its mean. This follows from linearity, which does not allow for a stronger form of relativity.

5.2 Neutrality: general aggregation of types

To generalize Theorem 5 for $W \in \mathcal{W}^{OEN}$, that is for more general aggregation of types, the following lemma will be helpful.

Lemma 2. Let $W \in \mathcal{W}^{OEN}$. Then $W(X) = \psi(u_1, \dots, u_n)$, for $u_h = \sum_{i=1}^{N_h} U_h(X_i^h)$ and $U_h(X_i^h) = \sum_{j=1}^k a_h X_{ij}^h + b_h$. Additionally, $a_h \frac{\partial \psi}{\partial u_h} > a_{h+1} \frac{\partial \psi}{\partial u_{h+1}}$.

Proof. Take any $X \in D$, let Y be obtained from X by means of within-type PDT for some vector $\varepsilon > 0$ and assume l to be a type for which distributions X, Y differ (for individuals r, s). Slightly abusing notation, we write $W(X) = \psi(U_l(X_s^l) + U_l(X_r^l))$. From INWT we have $W(X) - W(Y) = 0$ hence $\psi(U_l(X_s^l) + U_l(X_r^l)) - \psi(U_l(Y_s^l) + U_l(Y_r^l)) = 0$. It is equivalent to:

$$\frac{\psi(U_l(X_s^l) + U_l(X_r^l)) - \psi(U_l(Y_s^l) + U_l(Y_r^l))}{U_l(X_s^l) + U_l(X_r^l) - U_l(Y_s^l) - U_l(Y_r^l)} \cdot \frac{U_l(X_s^l) + U_l(X_r^l) - U_l(Y_s^l) - U_l(Y_r^l)}{|\varepsilon|(X_s^l - X_r^l)} = 0$$

Taking $|\varepsilon| \rightarrow 0$ we obtain:

$$\frac{\partial \psi}{\partial u}(U_l(X_s^l) + U_l(X_r^l)) \left[\frac{\partial u}{\partial X}(X_s^l) - \frac{\partial u}{\partial X}(X_r^l) \right] = 0$$

Therefore, either ψ is constant or $\frac{\partial u}{\partial X}(X_s^l) = \frac{\partial u}{\partial X}(X_r^l)$, hence the derivative of the utility function is constant and U_l is affine.

Now, again take any $X \in D$ and let Y be PDTT of X , with some $\varepsilon > 0$, for types l, m and individuals r, s from those types respectively. Abusing notation in a similar way as before, we write $W(X) = \psi(U_l(X_r^l), U_m(X_s^m))$. From IABT

$$\psi(U_l(X_r^l), U_m(X_s^m)) < \psi(U_l(Y_r^l), U_m(Y_s^m))$$

Dividing both sides by $|\varepsilon|$ and taking $\varepsilon \rightarrow 0$, we obtain the demanded condition. □

Theorem 6. *Let $W \in \mathcal{W}^{OEN}$ with $W(X) \neq 0$ for any $X \in D$. Then we have the following set of results.*

1. I_W is an inequality measure.
2. I_W is not relative or weakly relative.

Proof. We begin by checking the axioms. Continuity is obvious. If Y is fully egalitarian, then $W(Y) = W(1Y^\mu)$, hence by MON $\delta(Y) = 1$ and $I_W(Y) = 0$. Finally, if Y is PDTT of X , then by IABT $W(Y) > W(X)$ and since PDTT does not change total sums in any dimension, $W(Y^\mu) = W(X^\mu)$. Thus $\delta(X) > \delta(Y)$ and $I_W(X) < I_W(Y)$. From this, I_W is an inequality measure.

To show that I_W is not weakly relative, let $W(X) = \ln(\prod_{h=1}^n (\sum_{i=1}^{N_h} \sum_{j=1}^k a_j^h X_{ij}^h)) + 2 \sum_{j=1}^k a_j^1 X_{ij}^1$ and let $n = 2, N_1^X = N_2^X = 1, j = 2$. Set $X_{11}^1 = X_{12}^1 = 1, X_{11}^2 = 1, X_{12}^2 = 3$. Also set $a_1^1 = a_2^1 = 2, a_1^2 = a_2^2 = 1$. Such W clearly satisfies conditions of Lemma 2, hence $W \in \mathcal{W}^{OEA}$. Obviously $(X^\mu)_{11}^1 = (X^\mu)_{11}^2 = 1, (X^\mu)_{12}^1 = (X^\mu)_{12}^2 = 2$. By simple calculation, $W(X) = \log(16) + 8$ and $W(\delta X^\mu) = \log(18\delta^2) + 12\delta$ hence $\delta \approx 0.71$. Set C as a diagonal matrix with $c_{11} = 2, c_{22} = 2$. Now $W(XC) = \log(64) + 16$ and $W(\delta X^\mu C) = \log(72\delta^2) + 24\delta$ hence $\delta \approx 0.69$. □

In Theorem 6 we are unable to give an explicit formula for inequality measures induced by class \mathcal{W}^{OEN} as it is too general. However, it is possible to study the properties of IOp measures for this class, and these are stated in Theorem 6.

5.3 Aversion

It is easy to see that in the case of \mathcal{W}^{AOEA} we would also be unable to derive a formula for the IOp measure, but in this case we can find a subclass of welfare functions for which such explicit formula can be given. This subclass will be defined by the following well-known property (Tsui 1995).

Ratio scale invariance (RSI). For all $X, Y \in D$ and diagonal matrix C we have $W(X) = W(Y) \iff W(XC) = W(YC)$.

As seen by the definition of RSI, the subclass in question will consist of exactly those $W \in \mathcal{W}^{AOEA}$ for which I_W is relative. Please note that for \mathcal{W}^{AOEA} , δ is uniquely defined, since there is no matrix $Y \neq X^\mu$ with equal total sums in each dimension, such that $W(Y) = W(X^\mu)$. We now state and prove the following theorem, which is the main result of this paper.

Theorem 7. *Let $W \in \mathcal{W}^{AOEA}$, such that W satisfies RSI. Then we have the following set of results.*

1. $I_W(X)$ is a relative inequality measure.
2. Utility functions U^h are of the form $a_h \prod_{j=1}^k (X_{ij}^h)^{r_j}$, $a_h < 0, r_j < 0$.

3. $I_W(X)$ is given by

$$I_W(X) = 1 - \left(\sum_{h=1}^n w_h \frac{U^h((X_\mu)_1^h)}{U^h((X^\mu)_1^h)} \right)^{\frac{1}{\sum_{j=1}^k r_j}}$$

where $w_h = \frac{\delta_h(X) N_h a_h}{\sum_{h=1}^n N_h a_h}$ for $\delta_h(X)$ of the form

$$\delta_h(X) = \left[\frac{1}{N_h} \sum_{i=1}^{N_h} \prod_{j=1}^k \left(\frac{X_{ij}^h}{(X_\mu)_{1j}^h} \right)^{r_j} \right]$$

where $r_j < 0$.

4. For a matrix of type-means X_μ , $\delta_h = 1$ for all h and I_W measures only inequality between types.

Proof. First, we prove that I_W is an inequality measure. Continuity is obvious. $I_W(X^\mu) = 0$, since for $\delta = 1$ $W(X^\mu) = W(\delta X^\mu)$ and by monotonicity of W , it is uniquely defined. Since PDT does not change the mean value of each dimension, $I_W(X) > I_W(Y)$ if Y is PDT of X is implied by IABT. Relativity easily follows from the fact that $W(X) = W(\delta X^\mu) \iff W(XC) = W(\delta X^\mu C)$.

Let us fix h . We now use theorem 6 in Tsui (1995) to obtain the expression for δ_h —that is, the equality index for a single type h , defined as $W(X^h) = W(\delta_h X_\mu^h)$. The class $W \in \mathcal{W}^{AOEA}$ satisfies conditions in theorem 6, so it follows that utility functions are of the form $U^h(X_i^h) = a_h \prod_{j=1}^k (X_{ij}^h)^{r_j}$ with $a_h < 0, r_j < 0$ or $\sum_{j=1}^k r_j \log(X_{ij}^h)$, but as the second form does not satisfy IABT, 2 follows automatically. For this class of utility functions, the obtained equality index for a single type is $\delta_h(X^h) = \left[\frac{1}{N_h} \sum_i \prod_j \left(\frac{X_{ij}^h}{(X_\mu^h)_j} \right)^{r_j} \right]^{\frac{1}{\sum r_j}}$.

Now, relaxing the assumption of a single type, by ADD we have $W(X^1, \dots, X^n) = W(\delta_1 X_\mu^1, \dots, \delta_n X_\mu^n)$, with δ_h as above. We write

$$\begin{aligned} W(\delta X^\mu) = W(X) &\iff - \sum_{h=1}^n N_h a_h \prod_{j=1}^k (\delta (X^\mu)_{1j}^h)^{r_j} = - \sum_{h=1}^n N_h a_h \prod_{j=1}^k (\delta_h (X_\mu)_{1j}^h)^{r_j} \iff \\ &\iff \delta^{\sum_{j=1}^k r_j} \sum_{h=1}^n N_h a_h \prod_{j=1}^k ((X^\mu)_{1j}^h)^{r_j} = \sum_{h=1}^n N_h a_h \delta_h^{\sum_{j=1}^k r_j} \prod_{j=1}^k (\delta_h (X_\mu)_{1j}^h)^{r_j} \iff \\ &\iff \delta = \left(\sum_{h=1}^n w_h \frac{U^h((X_\mu)_1^h)}{U^h((X^\mu)_1^h)} \right)^{\frac{1}{\sum_{j=1}^k r_j}} \end{aligned}$$

as in the statement of the theorem. □

Theorem 7 states that inequality indices related to the class of inequality-averse, both within and between types, welfare functions are weighted sums of normalized types' utilities; that is, $\frac{U^h((X_\mu)_1^h)}{U^h((X^\mu)_1^h)}$, where weights are Tsui's (1995) equality indices computed within type. Inequality between types, on the other hand, is preserved by the fact that better types (i.e. types that have higher average utility), obtain lower weight. Less formally, one can view this measure as composed of two components, one for the distribution of attributes within type and one for the distribution of utilities between types. When there is no within-type inequality (i.e. X_μ), then within-type Tsui equality index is equal to 1 and I_W records pure inequality between types (point 4). Welfare function $W \in \mathcal{W}^{AOEA}$ is ratio-scale invariant, averse to CIT, and sensitive to PDT. These properties are ensured by the form of utility function with negative parameters a_h, r_j . This is the result of Tsui (1995) and Aczél (1988) that we use.⁴ Parameters r_j are dimensions'

⁴ For functions W that increase in CIT, the form of utility function would be the same with different parameters, namely $a_h > 0, r_j > 0, \sum_j r_j \leq 1$.

weights. The higher r_j (less negative) the lower the degree of concavity in a given dimension and the lower inequality weight attached to this dimension. To see this, let us assume that $r_1 = \dots = r_k = r$. We obtain the following:

$$\frac{d}{dr} I_W(X) = \ln(\delta(X))\delta(X) \frac{1}{Nr} \left[\sum_h \sum_i \ln \left(\frac{U^h(X_i^h)}{U^h((X^\mu)^h)} \right) \frac{U^h(X_i^h)}{U^h((X^\mu)^h)} \right] \leq 0$$

The sign of this derivative depends negatively on $\sum_h \sum_i \left[\ln \left(\frac{U^h(X_i^h)}{U^h((X^\mu)^h)} \right) \frac{U^h(X_i^h)}{U^h((X^\mu)^h)} \right]$. Note that this expression is a definition of generalized entropy of vector of utilities $U^h(X_{ij}^h)$, $h = 1, \dots, n$, $i = 1, \dots, N_h$ with parameter $\alpha = 1$. In this special case, generalized entropy is simply the Theil T index. As the Theil T index is non-negative, we obtain that increase in r results in decrease of inequality. Please note that since $r < 0$, an increase in r means decreasing the dimensions' weights.

6 Inequality of opportunity measures and dominance conditions

In this section, we show how IOp measures are related to implementable dominance conditions. Since inequality measures are derived from welfare functions, their relationship to dominance criteria follows from the link between welfare functions and dominance conditions established in Theorems 1–4. The difference is that the scale factor needs to be removed; recall that in the standard case (Atkinson 1970), inequality and welfare rankings are the same if, for example, the mean income is fixed. Here we have similar results, corresponding to the different reward principles.

Theorem 8. *Let $W \in \mathcal{W}^{AOEN}$ with $W(X) \neq 0$ for any $X \in D$. Then we have the following set of results:*

1. $I_W(X) \leq I_W(Y) \iff X'_\mu \succeq_{CGLD} Y'_\mu$, where $X_{ij}^{th} = \frac{X_j^h}{W(X^\mu)}$, $Y_{ij}^{th} = \frac{Y_j^h}{W(Y^\mu)}$.
2. If for all j , $\sum_{h=1}^n \sum_{i=1}^{N_h} X_{ij}^h = \sum_{h=1}^n \sum_{i=1}^{N_h} Y_{ij}^h$ then $W(Y) \leq W(X) \iff I_W(Y) \geq I_W(X)$.

Proof. We need to show that $\forall W \in \mathcal{W}^{AOEN} \quad I_W(X) \leq I_W(Y) \iff X \succeq_{CLD} Y$. By Theorem 1 we have $W(X') \geq W(Y') \iff X'_\mu \succeq_{CGLD} Y'_\mu$ hence $\frac{W(X)}{W(X^\mu)} \geq \frac{W(Y)}{W(Y^\mu)} \iff X'_\mu \succeq_{CGLD} Y'_\mu$ and $I_W(X) \leq I_W(Y) \iff X'_\mu \succeq_{CGLD} Y'_\mu$. Since 2 is a special case of 1, it follows from it. □

Theorem 8 states that the ranking of distributions according to IOp measures related to the class of welfare functions \mathcal{W}^{AOEN} is equivalent to the ranking of distributions according to CGLD, which acts on normalized matrices that are perfectly equal within type—that is, which acts on normalized X_μ . Here, the normalization factor is the value of the highest welfare; that is, this is achieved for the most equal distribution X^μ . Please note that it would not suffice to normalize attributes by their means because, as stated in Theorem 5 (point 3), the inequality measure is only weakly relative—that is, it is *not* invariant when attributes are scaled by *different* factors such as different means. If, however, outcome matrices X and Y have equal total sums for each attribute, then there is no need for any normalization (Theorem 8, point 2). For the case of general aggregation of types we obtain the same result, which is stated in Theorem 9.

Theorem 9. *Let $W \in \mathcal{W}^{OEN}$ with $W(X) \neq 0$ for any $X \in D$. Then we have the following set of results:*

1. $I_W(X) \leq I_W(Y) \iff X'_\mu \succeq_{CGLD} Y'_\mu$, where $X_{ij}^{th} = \frac{X_j^h}{W(X^\mu)}$, $Y_{ij}^{th} = \frac{Y_j^h}{W(Y^\mu)}$.

Proof. Since $\mathcal{W}^{AOEN} \subset \mathcal{W}^{OEN}$ then $I_W(X) \leq I_W(Y) \implies X'_\mu \succeq_{CGLD} Y'_\mu$. To obtain a reverse implication, note that by Theorem 2, $X' \succeq_{CGLD} Y' \implies W(X') \geq W(Y')$ for all $W \in \mathcal{W}^{OEN}$. Since by construction

total sums in each dimension of X', Y' are equal, it means that $\delta(X') \geq \delta(Y')$, and therefore $I_W(X) \leq I_W(Y)$, for any W . \square

We now formulate a result which shows that inequality ranking induced by inequality-averse IOp measures is implemented by the ranking of Lorenz dominance on some modified outcome distributions.

Theorem 10. *Let $W \in \mathcal{W}^{AOEA}$, such that W satisfies RSI. Then we have the following set of results.*

1. $I_W(X) \leq I_W(Y) \iff XC^X \succeq_{LD(\mathcal{F})} YC^Y$, where the class \mathcal{F} is as in Theorem 4, namely, \mathcal{U}^{ICL} .
2. The class \mathcal{U}^{ICL} here is restricted to utility functions of the form defined in point 2 of Theorem 7, namely, $U_h = a_h \prod_{j=1}^k (X_{ij}^h)^{r_j}$, where $|a_h| > |a_{h+1}|$ when $-\frac{1}{N_h} \sum_{i=1}^{N_h} \prod_{j=1}^k (X_{ij}^h)^{r_j} < -\frac{1}{N_{h+1}} \sum_{i=1}^{N_{h+1}} \prod_{j=1}^k (X_{ij}^{h+1})^{r_j}$.
3. These utility functions act on values of the attributes normalized by attributes' means; that is, C^X, C^Y are diagonal matrices with $c_{jj}^X = \frac{N}{\sum_{h=1}^n \sum_{i=1}^{N_h} X_{ij}^h}$, $c_{jj}^Y = \frac{N}{\sum_{h=1}^n \sum_{i=1}^{N_h} Y_{ij}^h}$.

Proof. Note that by RSI $I_W(XC^X) = I_W(X)$. By definition of C^X, C^Y we have $X^\mu C^X = Y^\mu C^Y$ since both of those have all values equal to 1. Assume $XC^X \succeq_{LD(\mathcal{F})} YC^Y$. By Theorem 4 and RSI we have

$$W(\delta(X)X^\mu C^X) = W(XC^X) \geq W(YC^Y) = W(\delta(Y)Y^\mu C^Y).$$

Since $X^\mu C^X = Y^\mu C^Y$ we have $\delta(X) \geq \delta(Y)$ and therefore $I_W(X) \leq I_W(Y)$. Since every operation was invertible, it concludes the proof. \square

The inequality ranking induced by IOp indices which are related to the W^{AOEA} class of welfare functions is equivalent to Lorenz dominance with two modifications. First, Lorenz dominance applies to the class of utility functions that are increasing, type-concave, concave, and submodular (\mathcal{U}^{ICL}). This class is further restricted by RSI, which gives the form as in point 2. Higher types' weights $|a_h|$ (please note that a_h is negative) are assigned to types that have lower average utility $-\frac{1}{N_h} \sum_{i=1}^{N_h} \prod_{j=1}^k (X_{ij}^h)^{r_j}$ (i.e. U_h with $a_h = -1$). This ensures that type-concavity holds globally. Second, Lorenz dominance acts on vectors of attributes normalized by their means—that is, C^X is a normalization matrix with diagonal elements one over the attribute mean. Please note that because in the aversion scenario the distribution within type matters, the relevant distributions which are compared by the Lorenz criterion cannot be summarized by X_μ .

7 Empirical application

We will now illustrate the methods developed in the paper with an empirical application. To this end, we use the US National Longitudinal Study of Adolescent to Adult Health (Add Health) to evaluate inequality of opportunity in the USA for the case of three dimensions of individual outcomes: education, health, and income. Add Health is a four-wave panel study that focuses on health-related behaviours and their causes. Initial information was collected in 1994/95 on adolescents in grades 7–12 ($N = 20; 745$) drawing on a stratified sample of 80 high schools in the USA. The sampling was conducted so as to ensure a nationally representative sample of adolescents enrolled in grades 7–12 in 1994/95. In addition to in-depth interviews with adolescents, questionnaires were administered to school representatives, parents, and roughly 90,000 students of the sampled schools. Importantly, the survey data are linked to additional contextual data from other data sources, such as the Census of Population and Housing, the

Table 1: Summary statistics for outcome variables

	N	Mean	SD
Personal income (k\$)	12.041	33,994.609	41,619.066
BMI	12.119	70.857	7.610
Years of education	12.284	14.120	2.202

Notes: means and standard deviations are weighted to correct for sampling procedure and sample attrition through waves 3 and 4.

Source: authors' compilation based on data from the Add Health survey, waves 1–4.

School District Databook, or the Statistics of the US Bureau of the State Government Finances. In the two most recent waves ($N = 15; 170$ and $N = 15; 701$, respectively) all respondents observed in Wave 1 had achieved the age of consent.

For the multidimensional outcome variable, we use self-reported personal earnings, years of education, and BMI in year 2008. The sample includes individuals age 24–32. Summary statistics are provided in Table 1.

Circumstance variables are derived from Wave 1, when the vast majority of respondents was younger than 18 years. We consider eight binary variables: race (white), maternal education (at least some college), paternal education (at least some college), biological sex (male), parental smoking (at least one smoker in the household), the marital status of parents (single-parent household), ability during childhood (above median Picture Vocabulary Test (PVT)), and parental income (above median household income). Table 2 presents the summary statistics.

Table 2: Summary statistics for circumstance variables

	N	Mean	SD
Race: white	12.282	0.704	0.457
Female	12.288	0.494	0.500
Educ. mother: >HS	12.285	0.354	0.478
Educ. father: >HS	12.283	0.297	0.457
Smoker in household	10.657	0.531	0.499
No father in household	12.288	0.421	0.494
PVT above median	11.718	0.530	0.499
Gross household-income above median	9.403	0.507	0.500

Notes: means and standard deviations are weighted to correct for sampling procedure and sample attrition through waves 3 and 4.

Source: authors' compilation based on data from the Add Health survey, waves 1–4.

There is considerable sample attrition. Wave 1 starts with 20,000 individuals, and 12,000 of them were interviewed in Wave 4. The analysis uses a pre-configured set of sampling weights to correct for this. Furthermore, there is selective item non-response for certain variables, which brings down estimation sample size to 8,726 individuals. To correct for non-response for different outcomes, the sample is re-weighted with respect to the socio-demographic characteristics such as race, maternal education, region of residence, and biological sex. Hence, the results are representative for the US population of adolescents enrolled in grades 7–12 in year 1994/95. Table 3 contains evidence to this effect. Despite sample reductions, the characteristics of each estimation sample do not differ significantly from the initial Add Health sample.

With eight binary circumstance variables, the maximum number of types is 192 in the most finely grained partition. Type partitions of increased granularity are created by sequentially introducing circumstance variables in the following order: (1) race, (2) sex, (3) maternal education, (4) paternal education, (5) household income, (6) parental smoking, (7) whether parents split, (8) child PVT outcome (above or below median). We first check the requirements of the same means rankings, which is necessary for the interpretation of CGLD and thus I_W^{AOEN} measure. This condition is trivially satisfied for all models in which we consider a single outcome. It is also satisfied for models with only binary partition (i.e.

Table 3: Robustness to selective attrition (*t*-test)

	Estimation sample			Full sample			Difference
	<i>N</i>	Mean	SD	<i>N</i>	Mean	SD	<i>p</i> -value
Education mother	8,726	0.297	0.457	18,908	0.286	0.452	0.072
Household income	8,726	0.704	0.456	18,913	0.698	0.459	0.324
Race	8,726	2.498	0.927	18,924	2.495	0.925	0.784
Religion	8,726	0.494	0.500	18,922	0.491	0.500	0.663

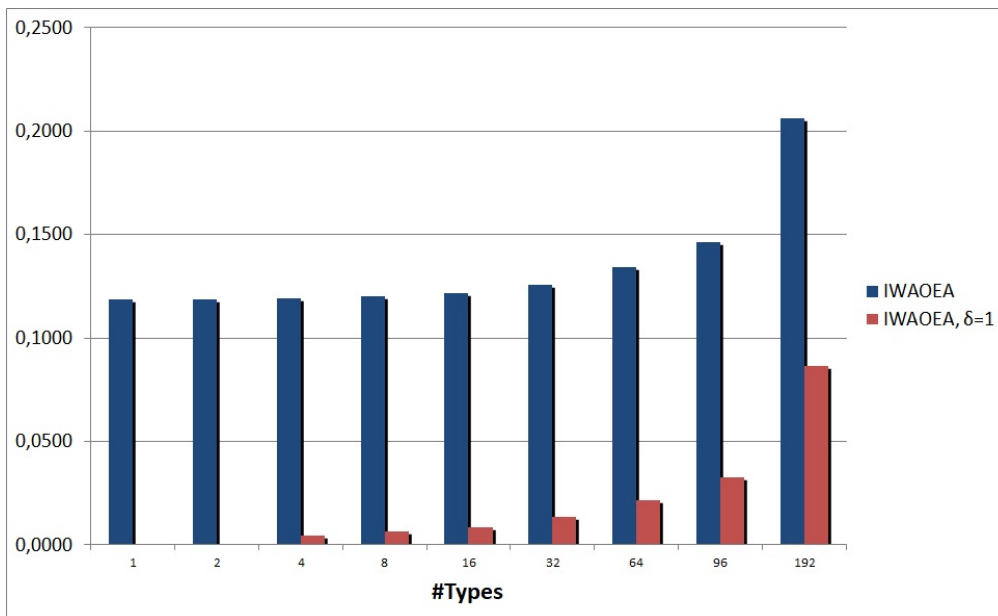
Notes: means and standard deviations are weighted to correct for sampling procedure and sample attrition through waves 3 and 4.

Source: authors' compilation based on data from the Add Health survey, waves 1–4.

only two types) for all outcome combinations (i.e. single outcomes, bidimensional outcomes, and joint three-dimensional outcome). However, already with a partition larger than two types, type rankings differ across dimensions. For this reason, in what follows we use the I_W^{AOEA} measure, which does not involve such restrictiveness.

Figure 1 shows the values of I_W^{AOEA} for three outcomes and varying type partitions, up to 192 types. For each partition, type weights are the following: for type *h* weight a_h is set as $196 - h$. We consider I_W^{AOEA} for the case when $\delta = 1$ and I_W^{AOEA} measures between-type inequality only, and I_W^{AOEA} without such a restriction. First, we note that the measure without restriction is non-zero when the whole sample is a type. This comes from the fact that I_W^{AOEA} takes into account inequality within type, which in this case exists. Second, I_W^{AOEA} increases with type partition, which replicates the well-known lower bound result by Ferreira and Gignoux (2011). This can also be viewed as an empirical confirmation that the measure is well defined. Third, aversion to within-type inequality is a strong driver of the results—more than half of IOp detected with 192 types is already present in the case where there is only one type. This is evident also when one compares the values of I_W^{AOEA} to the special case of $\delta = 1$ (red bars in Figure 1). IOp as measured by between-type inequality only is consistently less than half of I_W^{AOEA} (blue bars in Figure 1). Aversion to within-type inequality does matter for the evaluation of IOp.

Figure 1: Estimates of I_W^{AOEA} for joint outcomes

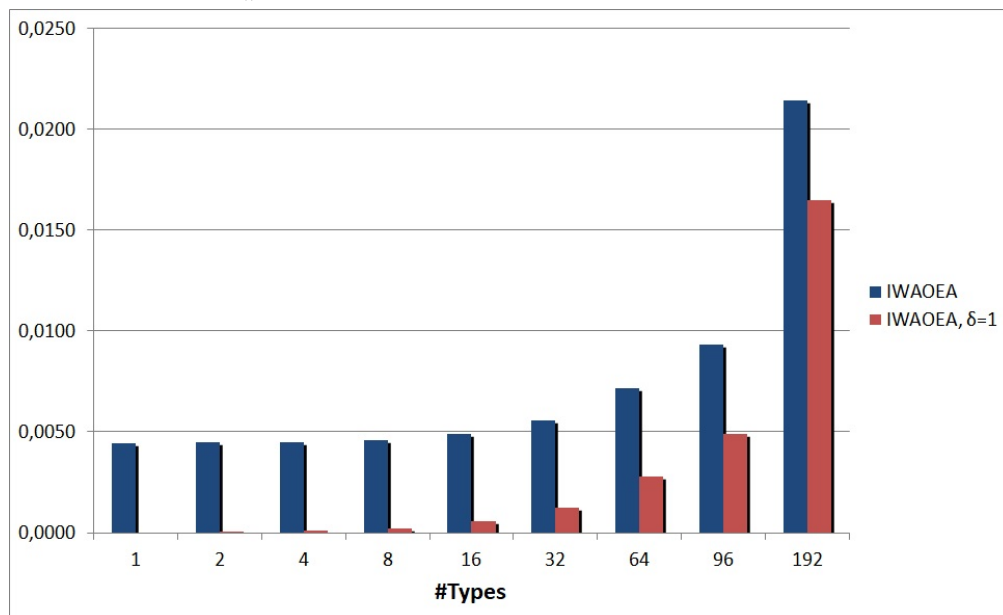


Source: authors' compilation based on data from the Add Health survey, waves 1–4.

We will now consider in more detail the issue of multidimensionality in the computation of IOp. Figures 2–4 show the values of I_W^{AOEA} computed for each outcome separately. These are cases of standard unidimensional IOp. IOp computed dimension by dimension differs significantly. Let us look closer at the case of maximum partition—that is, for 192 types. First, we consider the case of $\delta = 1$, so no within-

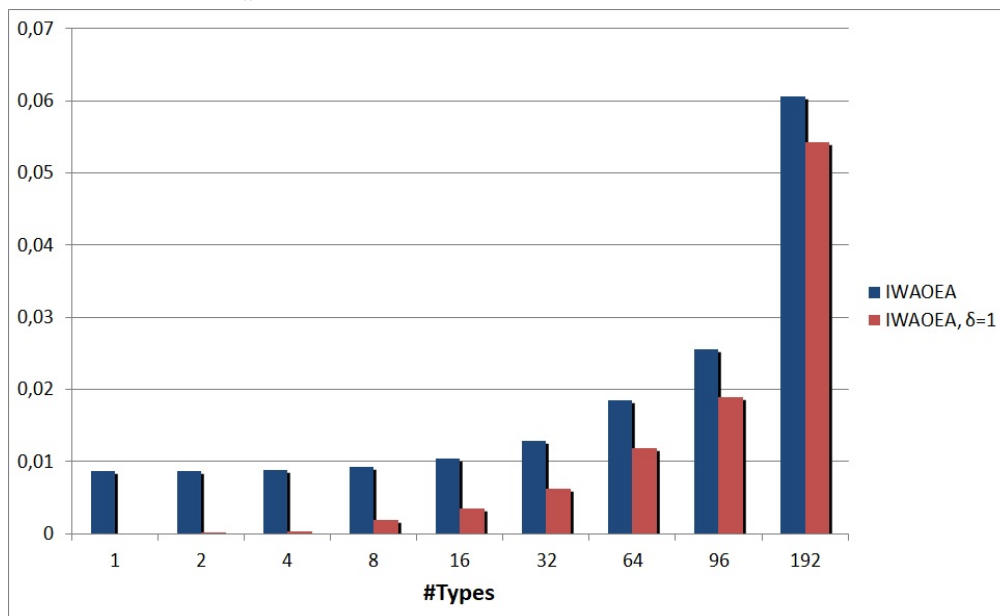
type inequality—comparison of red bars in Figures 1–4. The values of IOp for education and BMI are, respectively, 0.054 and 0.016, whereas for income it is 0.19. I_W^{AOEA} with $\delta = 1$ equals 0.086—in between the low IOp for education and BMI and the high IOp for income. This indicates the re-ranking of types with respect to mean values on each dimension, something we detected when checking the existence condition for CGLD. In other words, when dimensions are considered jointly, lower type position on one dimension is compensated via a higher position on another dimension, and a joint measure takes this into account. These differences are even more pronounced when additionally within-type inequality aversion is taken into account—comparison of blue bars in Figures 1–4. For 192 types, the values of IOp for education and BMI are, respectively, 0.06 and 0.021, whereas for income it is 0.47! Joint inequality I_W^{AOEA} equals 0.2. Here, this medium value of IOp for joint outcomes as compared to separate outcomes is—on top of types’ means re-rankings—caused by changes to within-type inequality. These changes are of two types, namely an outcome which is most unequal (income) is paired with outcomes which are distributed more equally (education and BMI) and/or high-income individuals are only to some extent best in education and BMI dimension. The former type of changes relates to the fact that when going from one dimension to three dimensions we take into account three marginal distributions. The latter type of changes relates to the fact that when going from one dimension to three dimensions, dependence structure between dimensions begins to matter.

Figure 2: Estimates of I_W^{AOEA} for BMI



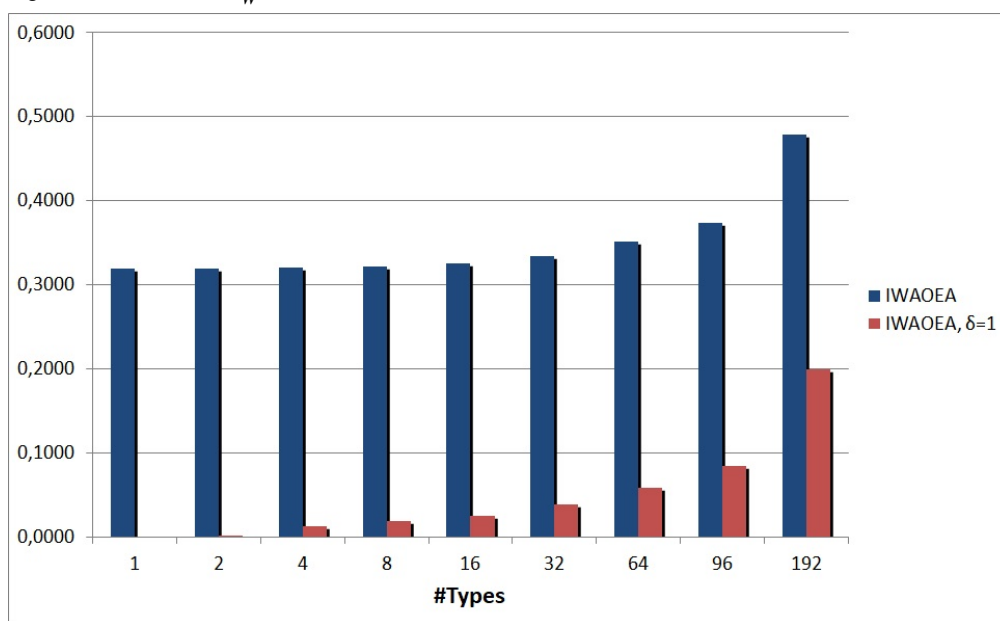
Source: authors’ compilation based on data from the Add Health survey, waves 1–4.

Figure 3: Estimates of I_W^{AOEA} for years of education



Source: authors' compilation based on data from the Add Health survey, waves 1–4.

Figure 4: Estimates of I_W^{AOEA} for income



Source: authors' compilation based on data from the Add Health survey, waves 1–4.

8 Concluding remarks

In this paper we develop a normative approach to the measurement of ex-ante IOp in a multidimensional setting—that is, when the individual outcome is represented by a multidimensional variable. We characterize axiomatically three classes of social welfare functions, all endorsing ex-ante compensation but each of them reflecting a specific reward principle: (1) utilitarian, (2) agnostic, and (3) averse. The first class is implemented via generalized Lorenz dominance applied to each attribute separately. The agnostic and inequality-averse classes are implemented by Lorenz ordering of type-aggregate utilities. The class of utility functions used in the inequality-averse scenario is submodular; hence it captures the dependence between attributes. We also develop normative inequality indices (Atkinson 1970; Kolm 1969; Sen 1973) for the classes of welfare functions and studied their properties. Finally, we propose

an empirical application of the methods developed in the paper: by using the Add Health survey in the USA we evaluate IOp in the USA for the case of three dimensions of individual outcomes: education, health, and income.

There are numerous theoretical extensions of this framework and this paper should be considered a starting point in the topic of extending EOp theory to a multidimensional setting. First, in this paper we focus on the ex-ante approach. A development of the present paper will study the multidimensional extension of the ex-post approach in all its variants (see Fleurbaey 2008; Fleurbaey et al. 2017; Peragine 2004; Roemer 1998). Second, we work with fixed population distributions. A natural extension is to allow for both changes in X and the partitions of N into types (see, e.g. Peragine 2004). Third, in the multidimensional case some attributes may not be transferable. It would be interesting to extend our results in this direction along the lines of Bosmans et al. (2009) and Muller and Trannoy (2012). These will be the subject of future research.

References

- Aczél, J. (1988). *A Short Course on Functional Equations*. Dordrecht: D. Reidel.
- 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(3): 244–63.
- Atkinson, A.B., and F. Bourguignon (1982). ‘The Comparison of Multi-Dimensioned Distributions of Economic Status’. *Review of Economic Studies*, 49(2): 183–201.
- Boland, P., and F. Proschan (1988). ‘Multivariate Arrangement Increasing Functions with Applications in Probability and Statistics’. *Journal of Multivariate Analysis*, 25: 286–98.
- Bosmans, K., L. Lauwers, and E. Ooghe (2009). ‘A Consistent Multidimensional Pigou–Dalton Transfer Principle’. *Journal of Economic Theory* 144: 1358–71.
- Bourguignon, F., F.H.G. Ferreira, and M. Menendez (2007). ‘Inequality of Opportunity in Brazil’. *Review of Income and Wealth*, 53(4): 585–618.
- Cohen, G.A. (1989). ‘On the Currency of Egalitarian Justice’. *Ethics* 99: 906–44.
- Dardanoni, V. (1992). ‘On Multidimensional Inequality Measurement’. In C. Dagum and A. Lemmi (eds), *Income Distribution, Social Welfare, Inequality and Poverty*. Stamford, CT: JAI Press.
- Dworkin, R. (1981). ‘What Is Equality? Part 2: Equality of Resources’. *Philosophy & Public Policy*, 4(10): 283–345.
- Ferreira, F.H.G., and J. Gignoux (2010). ‘Inequality of Opportunity for Education: Turkey’. In R. Kanbur and M. Spence (eds), *Equity and Growth in a Globalizing World*. Washington, DC: World Bank.
- Ferreira, F.H.G., and J. Gignoux (2011). ‘The Measurement of Inequality of Opportunity: Theory and an Application to Latin America’. *Review of Income and Wealth*, 57(4): 622–57.
- Ferreira, F.H.G., and V. Peragine (2015). ‘Equality of Opportunity: Theory and Evidence’. Policy Research Working Paper 7217. Washington, DC: World Bank.

- Fleurbaey, M. (2008). *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press.
- Fleurbaey, M., and V. Peragine (2013). 'Ex Ante versus Ex Post Equality of Opportunity'. *Economica*, 80(317): 118–30.
- Fleurbaey, M., V. Peragine, and X. Ramos (2017). 'Ex Post Inequality of Opportunity Comparisons'. *Social Choice and Welfare*, 49(3): 577–603.
- Gajdos, T., and J.A. Weymark (2005). 'Multidimensional Generalized Gini Indices'. *Economic Theory*, 26(3): 471–96.
- Idler, E.L., and S.V. Kasl (1995). 'Self-Ratings of Health: Do They Also Predict Change in Functional Ability?' *Journals of Gerontology: Series B*, 50B: S344–53.
- Kolm, S.-C. (1969). 'The Optimal Production of Social Justice'. In J. Margolis and H. Guitton (eds), *Public Economics*. New York: Macmillan.
- Kolm, S.-C. (1977). 'Multidimensional Egalitarianisms'. *Quarterly Journal of Economics*, 91(1): 1–13.
- Lefranc, A., N. Pistolesi, and A. Trannoy (2008). 'Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?' *Review of Income and Wealth*, 54(4): 513–46.
- Li Donni, P., V. Peragine, and G. Pignataro (2011). 'Measuring Equity in Health: A Normative Decomposition'. HERC Working Paper 11/06. York: University of York.
- Maasoumi, E. (1986). 'The Measurement and Decomposition of Multi-Dimensional Inequality'. *Econometrica*, 54(4): 991–97.
- Muller, Ch., and A. Trannoy (2012). 'Multidimensional Inequality Comparisons: A Compensation Perspective'. *Journal of Economic Theory* 147: 1427—49
- Paes de Barros, R., F.H.G. Ferreira, J.R. Molinas Vega, and J. Saavedra Chanduvi (2009). *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: World Bank.
- Peragine, V. (2002). 'Opportunity Egalitarianism and Income Inequality: The Rank-Dependent Approach'. *Mathematical Social Sciences*, 44(1): 45–64.
- Peragine, V. (2004). 'Ranking Income Distributions According to Equality of Opportunity'. *Journal of Economic Inequality*, 2(1): 11–30.
- Peragine, V., and L. Serlenga (2008). 'Equality of Opportunity for Higher Education in Italy'. *Research in Economic Inequality*, 16: 67–97.
- Ramos, X., and D. Van de Gaer (2016). 'Empirical Approaches to Inequality of Opportunity: Principles, Measures, and Evidence'. *Journal of Economic Surveys*, 30(5): 855–83.
- 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*, 22(2): 146–66.
- Roemer, J. (1998). *Equality of Opportunity*, Cambridge, MA: Harvard University Press.

- Roemer, J.E., and A. Trannoy (2015). 'Equality of Opportunity'. In A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, vol. 2A. Amsterdam: Elsevier.
- Rosa-Dias, P.R. (2009). 'Inequality of Opportunity in Health: Evidence from a UK Cohort Study'. *Health Economics*, 18(9): 1057–74.
- Sen, A.K. (1973). 'On Economic Inequality'. Oxford: Clarendon Press.
- Son, H.H. (2013). 'Inequality of Human Opportunities in Developing Asia'. *Asian Development Review*, 30(2): 110–30.
- Trannoy, A., S. Tubeuf, F. Jusot, and M. Devaux (2010). 'Inequality of Opportunities in Health in France: a First Pass'. *Health Economics*, 19(8): 921–38.
- Tsui, K.-Y. (1995). 'Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson–Kolm–Sen Approach'. *Journal of Economic Theory*, 67(1): 251–65.
- Tsui, K.-Y. (1999). 'Multidimensional Inequality and Multidimensional Generalized Entropy Measures: An Axiomatic Derivation'. *Social Choice and Welfare*, 16(1): 145–57.
- Van de Gaer, D. (1993). 'Equality of Opportunity and Investment in Human Capital'. PhD Dissertation. Leuven: Katholieke Universiteit Leuven.
- Yalonetzky, G. (2012). 'A Dissimilarity Index of Multidimensional Inequality of Opportunity'. *Journal of Economic Inequality*, 10(3): 343–73.