Volume 39, Issue 2, February 2011 ISSN 0305-750X

ELSEVIER

WORLD DEVELOPMENT

The multi-disciplinary international journal devoted to the study and promotion of world development

CONTENTS

SPECIAL ISSUE
ETHNICITY AND ETHNIC STRIFE
Guest Editors
Ravi Kanbur, Prem Kumar Rajaram and Ashutosh Varshney

(*continued on outside back cover*)

Indexed/Abstracted in: *British Humanities Index, CAB International, Current Contents, Geographical Abstracts, International Development Abstracts, Journal of Economic Literature, Management Contents, PAIS Bulletin, Sociological Abstracts, Social & Behavioral Sciences, Social Science Citation Index. Also covered in the abstracts and citation database SCOPUS®. Full text available on ScienceDirect®*

ELSEVIER
www.elsevier.com/locate/worlddev

World Development Vol. 39, No. 2, pp. 159–173, 2011

0305-750X/$ - see front matter

doi:10.1016/j.worlddev.2009.11.032

Inequalities between Groups: Theory and Empirics

ARJUN JAYADEV
University of Massachusetts, MA, USA

and

SANJAY G. REDDY *
The New School for Social Research, NY, USA

Summary. — Inequality between identity groups has long been thought of as an important contributor to social unrest and violence as well as being important in assessing the justice of societies. Yet, the measurement of the ways in which such groups differ and are unequal remains underdeveloped. Accordingly, this paper introduces three distinct but interlinked concepts relating to inequality between groups which can be used in empirical estimation of group based inequality. We define and discuss the concepts of representational inequality, sequence inequality, and group inequality comparison. Representational inequality captures the extent to which a given level of attribute is shared between members of distinct groups, sequence inequality captures the extent to which groups are ordered hierarchically in their possession of the attribute and group inequality comparison captures the extent to which differences between groups account for the overall inequality of individuals. These concepts can be used to measure the degree of segregation, clustering, and polarization between groups. In order to illustrate the merit of these concepts and their joint application to understanding group based inequality we provide an example using Demographic and Health Surveys data for five societies. It may be seen that the choice of measures is greatly consequential in applied work.

Key words — identity, inequality, segregation, polarization, clustering, groups, discrimination, conflict

1. INTRODUCTION

> "If we want a particularly satisfactory measure of inequality or poverty, we cannot define it over the income space alone and have to supplement the income data by information about the social relations between people and about comparison groups...Economic data cannot be interpreted without the necessary sociological understanding... There is a long way to go still to make adequate social sense of economic measures."
>
> Amartya Sen (2006, p. 38)

How much of the confrontation and hostility observed in many conflicts between identity groups today can be explained by the existence of inter-group differences in access to resources or in outcomes? Over the last decade or so, there has been an increasing scholarly interest in attempting to assess the degree to which resources, deprivations or well-being are unequally distributed across subgroups of populations. These concerns have led to the development of an expanding literature which has identified and empirically examined such concepts as "horizontal inequality" (see Stewart, 2001), segregation, polarization, (Zhang & Kanbur, 2001) and related ideas concerning differences between groups (see Subramanian, 2009 for a thoughtful discussion of such measures).

Some have argued that recently developed measures are better predictors of ethnic violence and civil conflict than previously developed measures such as inter-personal inequality or ethnic fractionalization (see in particular Montalvo & Reynal-Querol, 2005; Østby, 2008).[1] The literature on measurement of polarization has, despite its recent growth, made surprisingly little use of the existing apparatus of inequality measurement. Indeed, the subject of interest is often the degree of unevenness or inequality in the possession of attributes between groups, which is a concern both distinct from *and* related to the focus of traditional inequality measurement. Typical measures used in the literature (e.g., differences in mean achievements of the poorest and the richest groups in society) are often *ad hoc* and miss out important details (in this case, the degree of variance in achievements *within* groups and the achievements of groups which possess intermediate achievements).

Our paper seeks to address some of the weaknesses of the recent literature by building upon existing principles of inequality measurement. We provide a suggestive taxonomy of different aspects of inter-group inequality and develop a class of inequality measures for each of these aspects. We then show that these can be drawn upon further to develop measures of conjoint concepts (in particular, polarization). Because we rely on the many advances which have been made in the inequality measurement literature over the last few decades, we are able to provide a very wide range of measures to use to assess such differences. For every existing standard inequality measure there is a member of our class. The class of measures we introduce builds upon the existing literature in contrast with recent attempts to measure inter-group differences.

* We would like to thank for their useful comments or suggestions: (without implicating them in errors and imperfections): James Boyce, Glenn Loury, Rahul Lahoti, Hwok-Aun Lee, Yona Rubinstein, Peter Skott, Rajiv Sethi, Joseph Stiglitz, S. Subramanian, Roberto Veneziani and other participants at seminars in Brown University, the Dept. of Economics at the University of Massachusetts at Amherst, the Jerome Levy Institute at Bard College, Queen Mary, University of London, and the Brooks World Poverty Institute at the University of Manchester. We would also like to thank two anonymous referees for useful comments. The contents of this paper overlap partially with Reddy and Jayadev (2009). Final revision accepted: November 12, 2009.

The rest of the paper can be outlined as follows. In Section 2, we develop our taxonomy, making clear the informational underpinning of our concepts. For the most part we dispense with the technical and axiomatic undergirding of these concepts, except where essential [2] and instead rely on a graphical intuitive explanation. In Section 3 we provide examples of the application of our measures, employing a software package which we introduce. We use our measures to identify the extent of divisions between identity groups in five societies using Demographic and Health Surveys. We show that these measures allow for a richer discussion of the patterns of asset distribution among groups than might be obtained from currently used measures. In Section 4 we discuss the interpretative issues involved in applying quantitative measures of inter-group differences, including our own. In Section 5 we present some conclusions.

2. A SIMPLE TAXONOMY OF GROUP BASED INEQUALITIES

A common underlying concern in analyses of inter-group differences is the degree to which distinct groups are systematically over-or under-represented in their possession of various attributes (levels of income or health, club membership or political office, *etc.*). In this paper we introduce the concept of *representational inequality* (*RI*) as a way to capture this concern. This concept describes the extent to which a given attribute (for instance, a level of income or health, or right or left handedness) is shared by members of distinct groups. It can be used to measure the degree of "segregation" of distinct identity groups in the attribute space. [3]

When individuals can be ordinally ranked in relation to an attribute (such as income or health but not right or left handedness) we may be interested not only in how segregated or separated each identity group is in terms of their achievements, but in some measure of their relative positions in the ranking. *sequence inequality* (*SI*), understood as the degree to which members of one group are placed higher in a given hierarchy than those from another, captures this concern. Such a concept provides an intuitive framework for understanding the degree of "clustering" of various identity groups in distinct sections of a hierarchy. [4]

When individuals' level of achievement can also be cardinally identified for an attribute (as for income but not for right or left handedness) the *distance* between groups' attribute levels may be of interest. The concept of *group inequality comparison (GIC*) as we use the term here is based on comparison between the level of inequality in a society in which every member of a given identity group possesses a representative income (e.g., the mean income of their group) and the level of overall interpersonal inequality. [5] Related ideas have been explored extensively in economics (see, e.g., Shorrocks, 1980; Zhang & Kanbur, 2001). Measures used in group inequality comparison include but are not limited to the additively decomposable inequality measures which are the focus of that literature.

We show that combining these concepts can provide the basis of an understanding of conjoint concepts of group differences such as "polarization." [6] The concepts we identify can be fruitfully combined to measure polarization, understood to involve the collection of like elements and the separation of such collections of like elements from one another. The combination of representational inequality with sequence inequality alone provides a measure of what might be termed "Ordinal Polarization" Combining group inequality comparison with these two can provide a richer index of Polarization applicable to the case in which the attribute is cardinally measurable as well.

The concept of polarization that we employ here is distinct from that developed in the preponderance of the existing literature in that it draws on information about the identity groups to which those who possess distinct attributes belong. In contrast, the existing frameworks generally employ a "collapsed" framework in which the level of the attribute (typically income) defines the identity group (Duclos, Esteban, & Ray, 2004; Esteban & Ray, 1994). In these frameworks, polarization of an income distribution is understood to involve "identification" between individuals possessing a certain level of income and "alienation" between those individuals and others possessing different incomes. In our framework, in contrast, polarization of an income distribution is understood to involve segregation of individuals belonging to different identity groups at distinct levels of income as well as the extent of separation in the income space of these groupings of individuals belonging to a given identity group from other groupings of individuals possessing a distinct identity.

One approach to evaluating inter-group differences is to construct a measure of overall group advantage or disadvantage for each group prior to assessing the differences in these overall measures. [7] Although there can be advantages to such an approach, it can obscure the diverse aspects of inter-group difference (by reducing inter-group differences to inequalities in a single dimension). We accordingly explicitly identify here three distinct concepts of inter-group difference which we term representational inequality, sequence inequality, and group inequality comparison and a fourth (Polarization) which builds upon them. For expositional clarity, we limit ourselves to a two group example. All four concepts, however, are perfectly well defined for multiple group cases as well. For a more formal presentation of this point, see Reddy and Jayadev (2009).

(a) *Representational inequality*

Let us begin with the most general situation in which there is a population consisting of well-defined pre-existing groups possessing attributes of interest (income, health, extent of political representation, and so on).

We define a situation of representational inequality as occurring when, for some attribute and some identity group, the proportion of the group possessing the attribute is either greater or lesser than the proportion of the group in the overall population. To fix ideas, let us assume that the population consists of 50% "whites" and 50% "blacks" and that the attribute of interest is income (falling in pre-defined brackets). Figure 1 depicts the situation in which there is no representational inequality in the society. The location of each bar on the horizontal axis represents an income bracket ordered from lowest to highest and the proportion of persons possessing that income from either group is represented through shading. At all levels of income in this example, blacks and whites are represented in equal proportion to their share of the population as a whole (i.e., one-half each). *Any* deviation from such equiproportionality leads to a situation of representational inequality. Such a situation is depicted in Figure 2, in which at certain levels of income whites and blacks comprise a larger proportion of the individuals possessing that level of income than they do in the population as a whole. The figure happens to depict a situation in which, whites are disproportionally represented at the upper income brackets while blacks are disproportionately represented at lower income brackets.

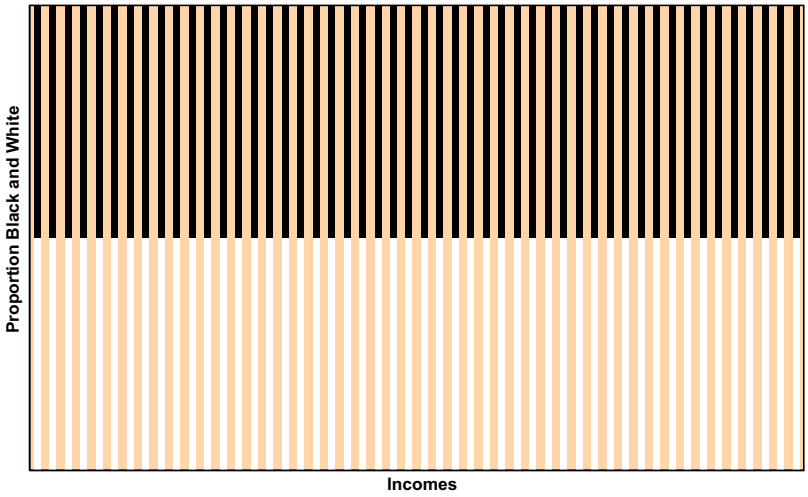

Figure 1. *Zero representational inequality.*

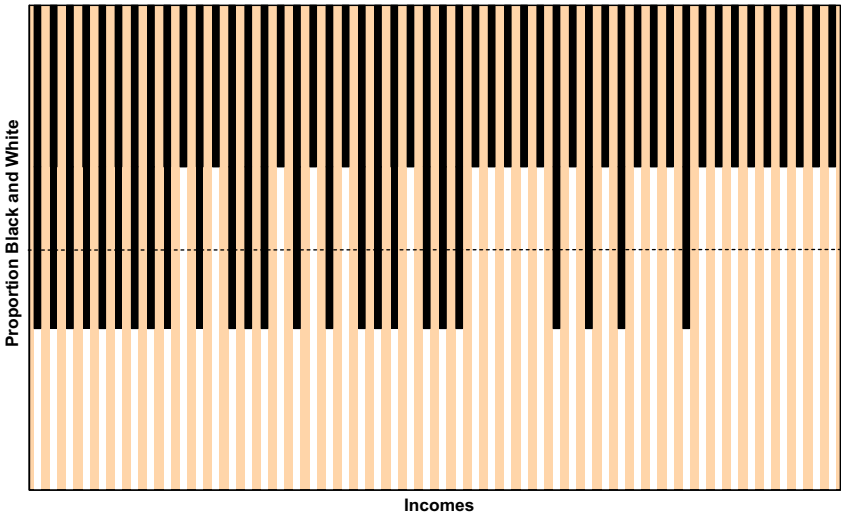

Figure 2. *Non-zero representational inequality.*

While the situation depicted in Figure 2 is one of representational inequality, both groups are represented at all the income brackets. In contrast, Figure 3 depicts a situation in which at each bracket there is *complete* segregation, in the sense that at each level of income there is one and only one identity group represented. It may be noted that while this is true, the overall salience of identity in determining the distribution of income may be called into question, since the incomes at which whites and blacks disproportionately appear are evenly interspersed. We depict this example in order to make sharp the distinction between segregation and clustering as we use the terms. The former refers to a situation in which those possessing a specific attribute (in this case income falling into a given income bracket) belong disproportionately to a particular group. The latter refers to a situation in which the attributes disproportionately possessed by members of a particular group are grouped together in a certain part of an attribute hierarchy (in this case the income spectrum). As such, representational inequality concerns the degree to which individuals from a given group are "localized" in their possession of attributes and not whether higher or lower levels of those attributes are systematically attached to members of a specific group.

How can one develop a way to measure the degree of representational inequality? Here we rely on some of the literature on Lorenz consistency (see, e.g., Shorrocks, 1983), which relates whether inequality measures respect a "transfer principle" to how they register an increase in inequality when the transfer brings about Lorenz dominance. This literature provides the basis for developing what we call a "representational inequality Lorenz curve" and a broad associated class of measures which obeys a relevant and connected transfer principle.

The *RI* Lorenz curve relates the cumulative population proportion of a given identity group, let us say whites, to the cumulative proportion of all others, when the proportions of those possessing a given attribute (i.e., fall into a given income bracket) and who are white are ordered from lowest to highest. This curve captures, for whites, the degree to which their representation in all income brackets differs from their representation in the population as a whole.

Specifically, in order to construct this curve, we first identify the population frequency of whites at every income bracket and then rank order these population frequencies from lowest to highest. For purposes of illustration, let us return to the concrete examples of Figures 1 and 3. In Figure 1, the fraction

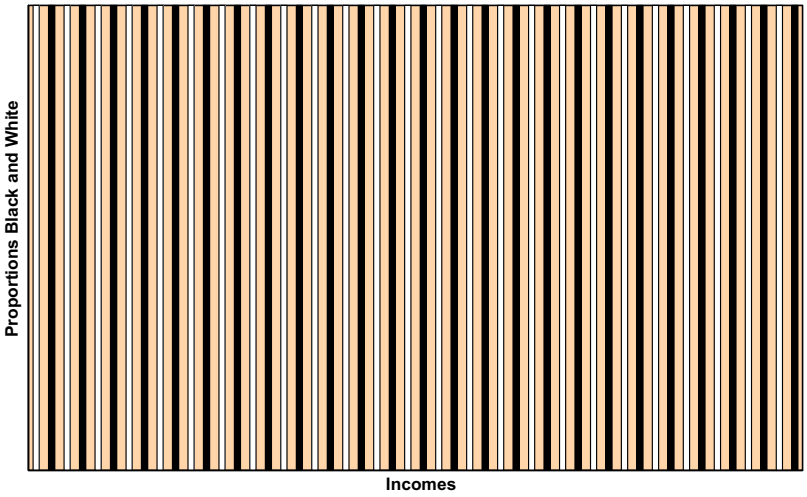

Figure 3. "*Complete*" *segregation.*

of the population at every income bracket which is white is one-half. We now trace out the cumulative proportion of whites as we add the proportions of blacks at different brackets where these brackets are ordered from that possessing the lowest white population frequency to that possessing the highest white population frequency. For every additional income bracket that is added, whites are added in the same population frequency as the frequency of blacks (1/2), so that the Lorenz curve hugs the 45° line. As such, the resulting representational inequality Lorenz curve depicts zero representational inequality. Applying a similar argument, the same would be true for the *RI* Lorenz Curve for blacks. In Figure 3, however, there are 25 brackets in which whites are represented with zero population frequency and 25 brackets in which they are over represented (i.e., representing 100% of the population in that income bracket). In this case, our ordering of the brackets will result in a series (0, 0, 0, 0, 0, ..., 1, 1, 1, 1, 1, 1). As such, as we add blacks, no whites are added until we exhaust all the income brackets in which blacks are represented (and correspondingly all of the blacks). As such, the representational

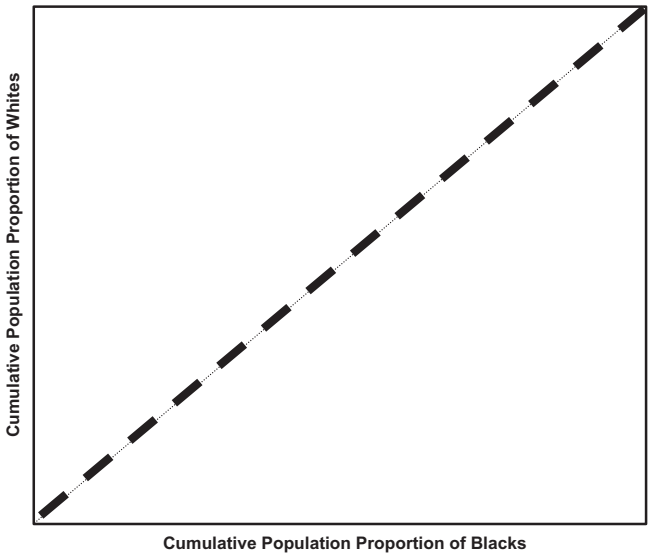

Figure 4a. *Zero representational inequality Lorenz curve.*

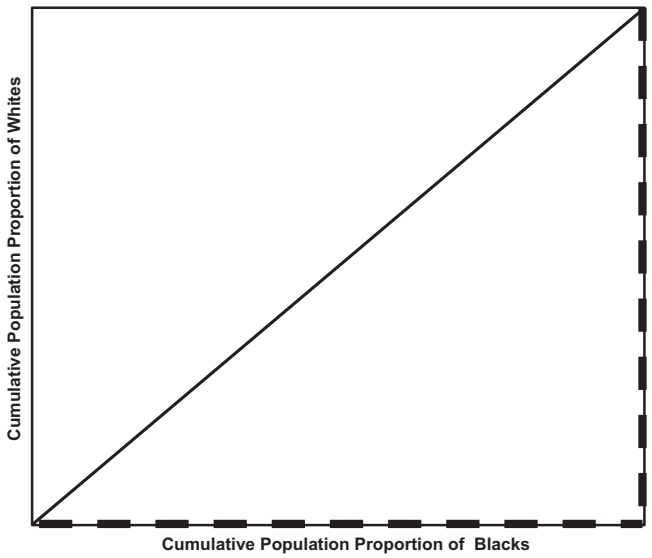

Figure 4b. *Maximal representational inequality Lorenz curve.*

inequality Lorenz curve is horizontal along the axis until it reaches the lower right hand side of the Lorenz box, when it rises vertically. As such, it depicts maximal representational inequality. By a corresponding argument, the same would be true for the *RI* Lorenz Curve for blacks. These two examples are given in Figures 4a and b.

What is the form of transfer which brings about dominance of *RI* Lorenz Curves? It is easy to see that by shifting blacks from where they are more heavily represented to where they are less heavily represented the population fraction of blacks and of whites can simultaneously be made more equiproportional. A transfer of this kind will bring about a new *RI* Lorenz Curve which dominates the previous one.[8] Inequality measures which respect such dominance are those which register a decrease when such a transfer takes place. The existing literature on inequality measures can be used to identify which measures possess this property. Indeed, all standard inequality measures obey the Pigou–Dalton transfer principle, and can correspondingly be applied here.

The concept of representational inequality clearly need not be restricted to a scenario in which the attribute is cardinally orderable. Thus, for example, we can apply the principle in a straightforward manner to unordered attributes such as location of residence or membership in distinct clubs or legislatures. If instead of income brackets, each bar in the figures referred to a distinct state legislature in a federal country, the figures we have discussed here would depict the degree of inequality in the political representation of groups.

(b) *Sequence inequality*

The distinction between "complete segregation" and "complete clustering" can be seen by comparing Figures 3 and 5. Figure 5 depicts the situation that results from a transfer of incomes such that all the whites move to the richer half of society while all the blacks move to the poorer half of society. This situation is one in which each sub-group is concentrated in a different part of the income distribution. Such a situation can plausibly be described as one of "complete clustering" of groups. In both cases, there is complete segregation and thus maximal representational inequality. However, in Figure 3, whether an individual is black or white provides very little information on his or her rank in society. By contrast, in Figure 5, whether an individual is black or white provides a great deal of information. One simple way to capture the distinction between Figures 3 and 5 is through the concept of sequence inequality, which can together with representational inequality capture the clustering of the income distribution. This concept is linked to the position in the overall societal ranking possessed by individuals belonging to distinct groups in the hierarchy.

An individual (weakly) rank-dominates another if that individual is ranked equal to or higher than the other in the possession of income. For any population partitioned into given identity groups, there are a fixed number of between-group pair-wise comparisons between individuals from different identity groups. Consider the example introduced in the previous section and assume that there is one individual in each income bracket, so that there are 25 whites and 25 blacks. In this case, there are 625 between group pair-wise comparisons. The share of the total number of such between group pair-wise comparisons involving a given group in which a member of the group rank-dominates a member of some other group is called its level of group rank dominance. For Figure 3, whites dominate in 325 between group pair-wise comparisons and their level of group rank dominance is therefore 325/625 or 0.52.[9] Correspondingly, blacks dominate in 300 pair-wise comparisons and their level of group rank dominance is 300/625 or 0.48. For Figure 5, by contrast, whites dominate in 625/625 between group pair-wise comparisons and their level of group rank dominance is 1. Correspondingly, blacks dominate in 0 between group pair-wise comparisons and their level of group rank dominance is 0. Group rank dominance is an indicator of the position the group occupies in the ordinal hierarchy of attribute levels. Another way to understand the difference between Figures 3 and 5 is simply that the average rank of the whites and the blacks is different. This is clearly a necessary condition for distinct groups to be clustered in different parts of the attribute space.[10] It is clear from this discussion that while Figures 3 and 5 depict two groups with equal representational inequality, since there is complete segregation in both cases. However, the two groups possess different levels of group rank dominance and average rank. In Figure 5, whites have 100% of the available instances of rank domination and higher average rank.

The level of inequality in different groups' rank sequence position (whether as measured by group rank dominance or by average rank) indicates the extent to which a population is clustered. We refer to this concept of inequality as sequence inequality (*SI*). Some reflection will suffice to show that this is an unambiguous criterion even when group sizes differ. In any situation sequence inequality is minimal when the groups are evenly interspersed or symmetrically placed around the median member(s).

How can one assess the level of overall sequence inequality in the society given that the group rank sequence position (GRSP) of each group has been determined? We may note that if the groups possess the same GRSP then there cannot be any sequence inequality as we have presented the term. This suggests that we can think of sequence inequality in relation to a transfer principle. In particular, an exchange of relative position between two individuals belonging to distinct groups which brings about an increase in GRSP for the group possessing the lower GRSP initially and a decrease in GRSP for the group possessing the higher GRSP initially (without changing the relating GRSP of the groups) must bring about a reduction in sequence inequality. For example, imagine that

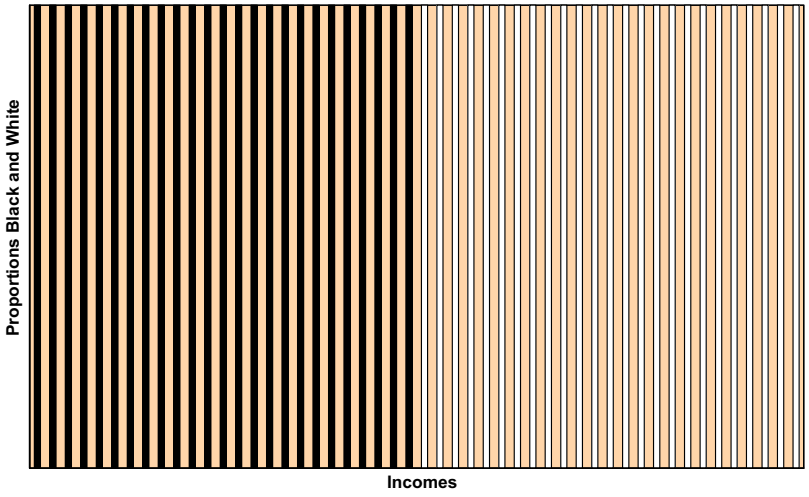

Figure 5. *Clustering.*

the richest black and the poorest white in Figure 5 interchanged incomes. This would result in whites having less than 100% of the instances of rank domination, and blacks correspondingly having more than 0% of the instances of rank domination and thereby reduce the level of overall sequence inequality.

We may say that a "progressive" transfer of GRSP must bring about a reduction in overall sequence inequality and that a "regressive" transfer must bring about an increase. It is easy to see that any standard inequality measure respecting the Pigou–Dalton transfer principle when applied to the GRSP of groups can serve as a measure of overall sequence inequality. For example, one can assign to each individual belonging to a group the GRSP score of that group and the inequality measure can be applied to the resulting synthetic population. The GRSP scores in the synthetic population can be depicted using an appropriate Lorenz curve.

To see this, consider Figures 3 and 5 again. In Figure 3, we have 25 individuals (blacks) with group rank dominance of .48 and 25 individuals (whites) with group rank dominance of .52, that is, a series (.48., .48, .48, ..., .52, .52, .52). Drawing a Lorenz curve for these individuals and assigning them these scores will result in a Lorenz curve which depicts near equality (Figure 6a). In Figure 5, we have 25 individuals (blacks) with group rank dominance of 0 and 25 individuals (whites) with group rank dominance of 1, that is, a series (.0, .0, 0, ..., 1, 1, 1). Drawing a Lorenz curve for these individuals and assigning them these scores will result in a Lorenz curve which depicts more inequality (Figure 6b). It can be observed that a series of exchanges of position between individuals which result in the regressive transfer of GRSP will bring about a shift from the first situation to the second.

Note here that the measure of sequence inequality is dependent on the relative population size. In particular, sequence inequality tends to its maximal value when a single member of a group dominates members of another single large group.

While sequence inequality and representational inequality are related, they are also distinct concepts. A simple example which makes this distinction transparent is provided in Figures 7a and 7b. In Figure 7a, both groups possess the same level of group rank dominance and average rank. The black group has

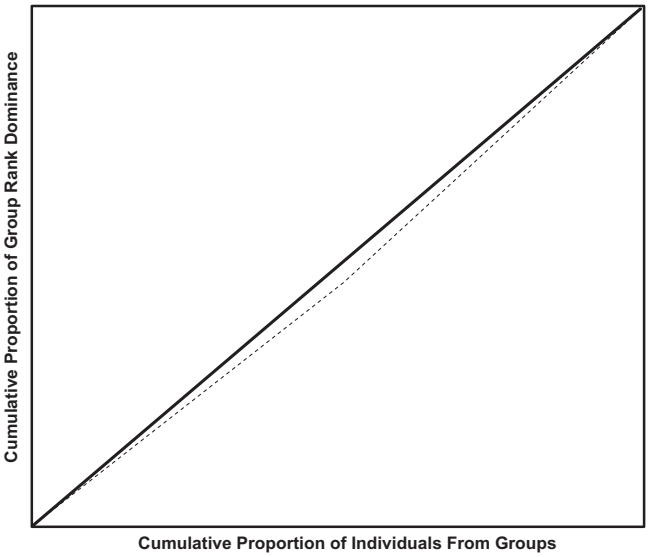

Figure 6a. *The GRSP Lorenz curve for Figure 3.*

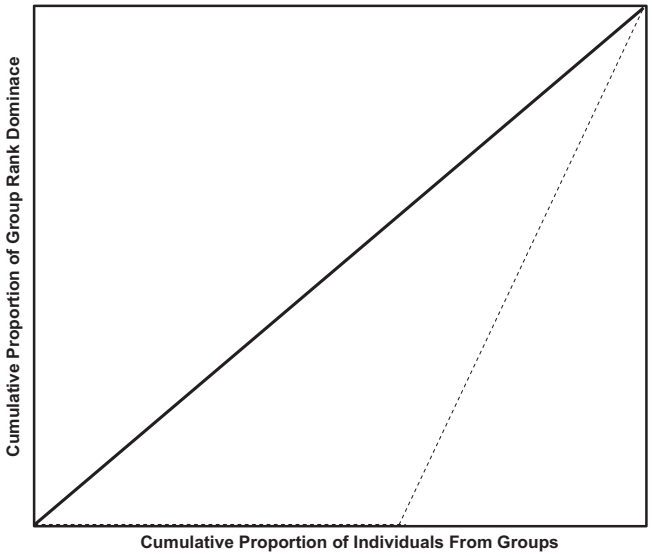

Figure 6b. *The GRSP Lorenz curve for Figure 5.*

two of the possible four instances of rank domination as does the white group, and their average rank is the same. Thus there is no sequence inequality between the groups. In Figure 7b, both groups again share equally in levels of group rank domination (both have two of the potential four instances once again) and have the same average rank. The situation once again is one in which there is no sequence inequality. However, in the first case there is complete representational inequality and in the second case there is zero representational inequality. In neither case is group membership always associated with higher rank, yet the cases differ in the degree to which income levels are shared by members of distinct groups.

(c) *Group inequality comparison*

Figure 5 depicts a situation of maximal representational inequality and maximal sequence inequality. It could perhaps be thought of as a situation of polarization in the sense that each group is concentrated at a given pole of the income distribution. However, both the situations depicted in Figure 5 and in Figure 8 are *identical* from the standpoints of representational inequality and sequence inequality since neither concept takes note of cardinal information, which alone accounts for the difference between the two situations described. To take account of cardinal information (for instance, concerning the distance between the distinct clusters), it is necessary to introduce an additional concept.

A common way to account for such information is to take note of the distance between the means of distinct sub-populations, for example, by using measures of inequality between group means. For a fuller discussion of the appeals of this approach, see Reddy and Jayadev (2009). However, such an approach ignores relevant information on *within* group inequality. Consider a two-group society in which all members of each group originally, respectively, possess the mean incomes of their groups. Suppose that both groups experience within-group regressive transfers leading to intra-group inequality. The extent of inequality in the society must be judged to have increased if the measure of inequality employed obeys the Pigou–Dalton Transfer Principle (ensuring that such transfers between persons are deemed to increase overall inequality). However, between-group inequality (understood

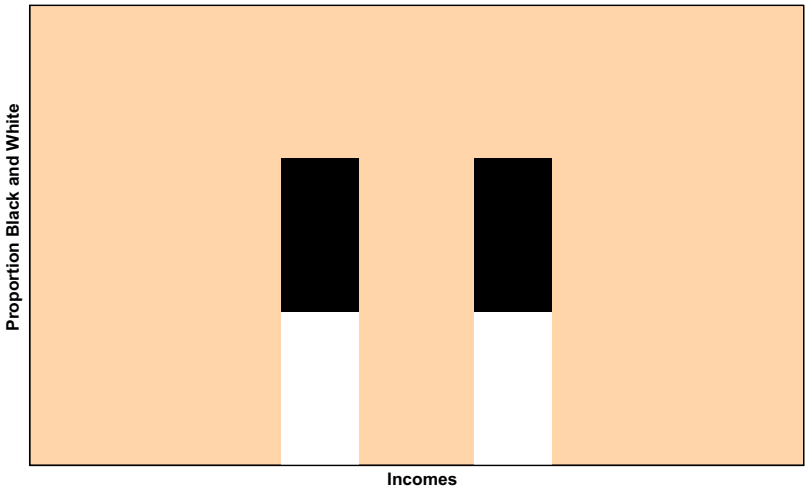

Figure 7a. *Perfect sequence equality with perfect Representational Equality.*

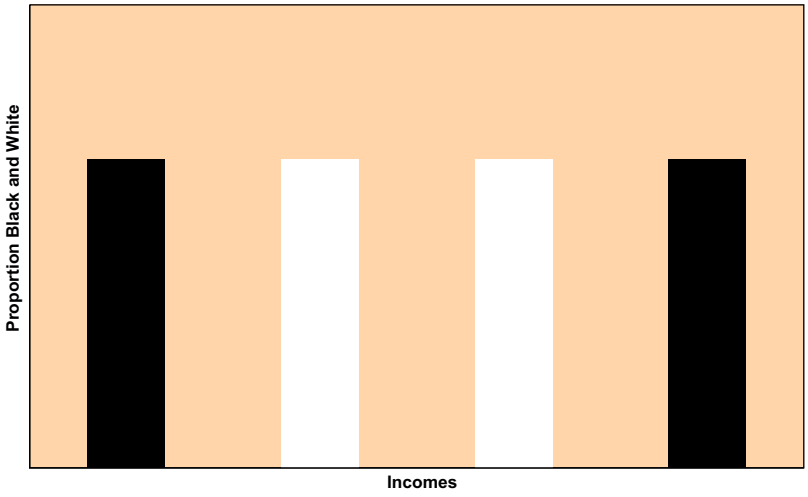

Figure 7b. *Perfect sequence equality with complete segregation.*

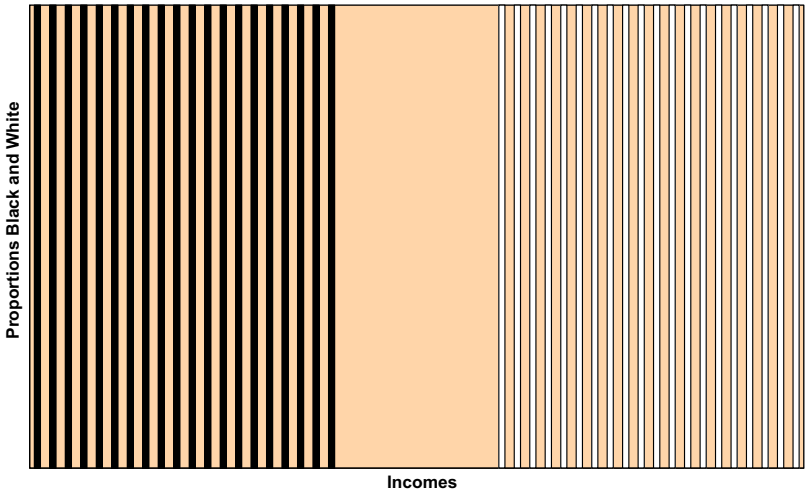

Figure 8. *Polarization.*

in terms of inequality between mean incomes of groups) is unchanged. The role played by between-group inequality in generating overall interpersonal inequality must be deemed to have decreased and consequently the importance of

inter-group differences in the determination of overall social inequality must also be judged to have decreased.

An approach to inter-group inequalities which is based on between-group inequalities in isolation rather than on the contribution of between-group inequality to overall inter-personal inequality may fail to contrast situations that might be distinguished. The concept of inter-group difference is a complex one, which is not always plausible to reduce to differences between means. Consider Figure 9 which depicts a two group society in which all members of each group originally possess mean income A and B, respectively. Suppose that both groups now experience within-group transfers which increase inequality and their distributions are now depicted by densities A′ and B′, respectively. Assume further that the transfers are such that the span between the means is Δ and the span between the richest and poorest members of each group is also Δ. One may consider inter-group differences to have become less significant after the transfer since no member of the richer group is further away from some member of the poorer group than before the transfer, and all but the very richest member of the richer group is closer to some member of the poorer group.

The concept of group inequality comparison we apply in this paper is based on comparing the inequality arising when each member of a group is assigned the same "representative" income and overall interpersonal inequality. The representative income used in the former case need not be the mean and could be the median or another representative measure.[11] If between-group inequality is defined as the inequality of the population that arises when every member of the population is assigned the representative level of attribute of the group to which they belong, then the ratio of between-group inequality to total interpersonal inequality can serve as an adequate measure of group inequality comparison. This approach is based on the comparison of counterfactuals and any inequality measure can be used in such comparison.

(d) *Combining concepts: polarization*

We have introduced above three concepts relating to inter-group inequalities: representational inequality, sequence inequality and group inequality comparison. How are these concepts related to polarization? Polarization is a concept which has been used in many different ways in the literature, for example, to mean the absence of "middleness" in a distribution (Wolfson, 1994), the distance between the average achievements of the groups weighted by the size of the groups in such a way that the increment in weight attached to a group due to an increment in its size is larger when the group is larger (Østby, 2008) and the "distance from" a distribution consisting of two equally sized groups (Montalvo & Reynal-Querol, 2005). These approaches do not always identify the groups to which individuals belong independently from the attributes which they possess.

If we understand the level of polarization of a distribution in terms of the extent of differences between identity groups in the possession of an attribute, it becomes clear that each one of the concepts of inter-group inequality (*RI, SI*, and *GIC*) is *itself* a measure of polarization. In the case of unranked attributes (e.g., if the figures used referred to membership of blacks and whites in clubs), *RI* alone is sufficient to establish polarization of society. In the case where attributes are rankable but not necessarily cardinally measurable (e.g., health achievements), *RI* and *SI* together may be combined to establish polarization. Finally, in the case where attributes are cardinally measurable (e.g., income), *RI, SI*, and *GIC* together may be combined to establish polarization.

The relative ranking of the situations depicted in Figures 3, 5 and 8 differs according to the extent of polarization and depends on the expansiveness of the approach used. In particular, all the figures depict maximal polarization as judged according to *RI*, whereas Figures 5 and 8 depict maximal polarization according to both *RI* and *SI*, and Figure 8 depicts a higher degree of polarization than does Figure 5 according to *GIC* (taking the figures to possess the same income scale on the horizontal axis).

The fact that our judgments regarding the polarization of society may depend on more than one concept suggests the value of combining measures of inter-group differences to construct orderings of social situations according to the extent of their polarization.[12] Furthermore, assessing polarization may require the combination of concepts. We approach this issue once again with some simple graphical intuitions. It is easy to see from our example that *RI* alone is insufficient to establish polarization when attributes are rankable (it would, e.g., fail to differentiate between Figures 3, 5 and 8). *SI* alone is also insufficient to establish polarization when the attribute is cardinally measurable (it fails to distinguish between Figures. 5 and 8).

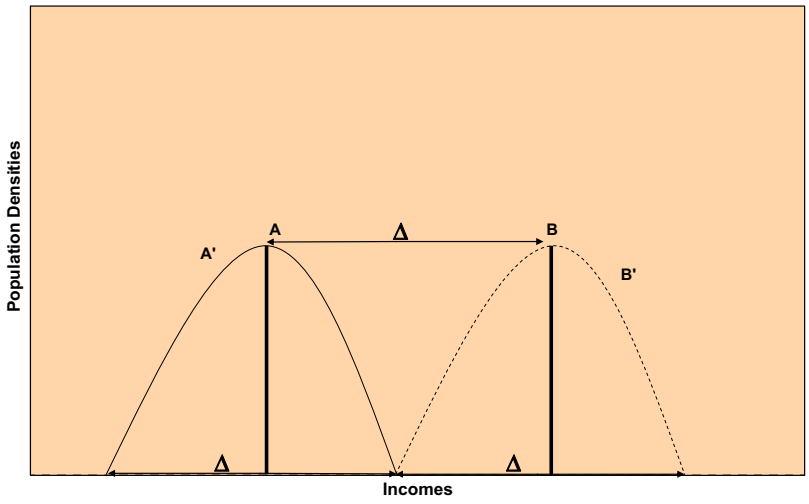

Figure 9. *Group inequality contribution versus inequality between means.*

Can *GIC* serve as a stand alone measure of polarization? A variant of group inequality comparison has been proposed for this purpose (Zhang & Kanbur, 2001). However, such a measure, while attractive in its simplicity can violate some intuitions. Consider Figure 10 in which two completely segregated and clustered groups A and B experience within-group progressive transfers which reduce within-group inequality. Further, suppose that they also experience a reduction of between-group inequality through progressive transfers between the members of the two groups in such a way that the ratio of between-group inequality to overall inequality remain unchanged and the groups (whose densities are now depicted by A′ and B′) overlap. If we utilize group inequality comparison alone as our measure of polarization, a social configuration with A and B is viewed as being exactly as polarized as a situation with A′ and B′, which seems to conflict with our intuitions. If, however, we combine it with some measure of sequence inequality and/or representational inequality (both of which are lower when the groups overlap) the first situation is unambiguously more polarized than the second.

Figure 10 suggests that representational inequality and sequence inequality capture the degree of "overlap" between groups, which it may be argued is central to our intuitive understanding of polarization, and that *GIC* by itself does not. A reasonable question that may arise is whether *SI* and *GIC* taken together are sufficient to capture such intuitions. Figures 7a and 7b suggest that *RI* has a crucial role to play. In both Figures 7a and 7b, there is zero sequence inequality and *GIC* could be assumed to be the same, but in the former there is zero representational inequality while in the latter there is maximal representational inequality. As such, one might argue that inter group differences are more salient in the situation described in Figure 7b than in Figure 7a.

Polarization, by this reasoning could be measured by aggregating the three concepts concerning group differences which we have defined. The range of polarization measures which could be used is very wide indeed since any such measure could involve any form of aggregation of a three-tuple (*RI*, *SI*, and *GIC*), and in turn each element of this three-tuple could be defined in various ways. Further, any measure of polarization which is positively responsive to all three will only be maximized in a situation where all three are maximized. One obvious and attractive measure would be a multiplicative form—that is, $P = RI * SI * GIC$ (or a monotonic transformation thereof). [13] We specialize to this case in the empirical work which follows purely for expositional simplicity.

Each of the three constituent elements of polarization can be measured using any Lorenz consistent measure. Not all Lorenz consistent inequality measures have a finite range and therefore are not readily interpretable. In particular, all members of the generalized entropy class of inequality measures which satisfy the desirable properties of replication invariance and the transfer principle have an infinite range (see Shorrocks, 1980). The generalized entropy class of measures nevertheless has a particular appeal for the measurement of *GIC* in that the ratio of between-group inequality to overall inequality is the fraction of overall inequality that is "accounted for" by between-group inequality. [14] This said, however, inequality is measured, an index of *GIC* based on the ratio of inequality when persons are assigned the representative income of the group to which they belong to overall interpersonal inequality is bounded between zero and one.

It may seem esthetically appealing to apply the same inequality measure in calculating all three of the component parts of the polarization measure. However, there are tradeoffs involved between the pursuit of this esthetic goal and the pursuit of other goals. In particular, if a non-normalizable measure of inequality (such as a member of the generalized entropy class) is used to assess all three component forms of inter-group inequality then the resulting polarization measure will also not be normalizable. It follows that if the same inequality measure is to be applied uniformly to calculate all of the components of the polarization measure then the use of an additively decomposable measure (of the kind proposed by Shorrocks, 1980) to calculate *GIC* cannot be reconciled with the goal that the overall polarization index should have a finite range.

This creates a choice for the analyst who desires that the polarization index be normalizable. Such an analyst may either adopt a "non-uniform" approach in which distinct inequality measures are used to calculate *GIC* on the one hand and *RI* and *SI* on the other, or alternatively may adopt a uniform approach in which the inequality measure used is not similarly additively decomposable but is normalized in the desired way.

Polarization is maximized in a particular configuration. Note that a social configuration in which *RI*, *SI*, and *GIC* simultaneously approach their respective maximum values is

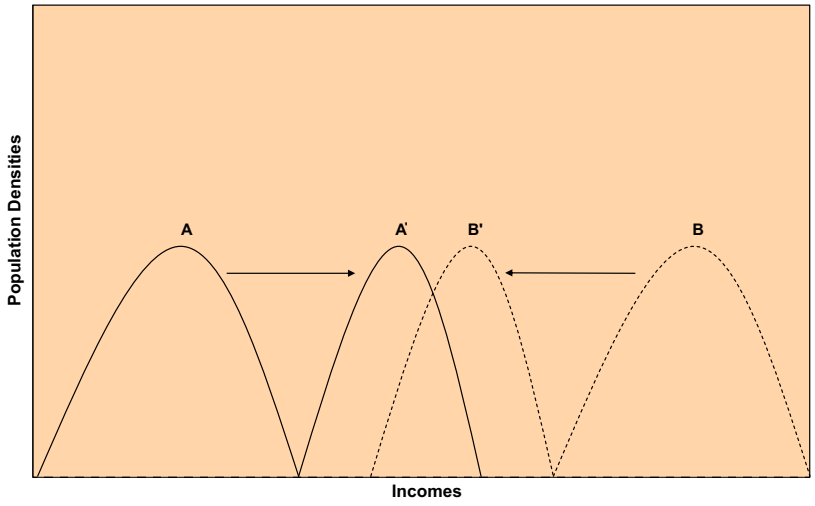

Figure 10. *Group inequality contribution alone is an incomplete measure of polarization.*

one in which polarization (understood as a positively responsive function of these three elements) approaches its maximum. Does such a social configuration exist? *RI* is maximized when there is complete segregation of groups in the achievement space. Assuming this condition is satisfied, what additional conditions are required to maximize *SI*? It is easily seen that the maximum value of *SI* is approached in the limit in which the social configuration is such that one group is as large as possible relative to all others and is disadvantaged relative to all others (specifically in the sense that every member of the disadvantaged group possesses lower achievements than every member of every other group). Assuming this condition too is satisfied, *GIC* in turn is maximized when the within-group inequalities in the population are as small as possible. This will be attained when there are no intra-group differences in achievements. Any social configuration which jointly satisfies these properties will suffice to approach the maximum level of polarization. One such example is provided in Figure 11.

3. NUMERICAL AND EMPIRICAL EXAMPLES

An empirical examination which involves these four concepts can, as we have noted, be implemented through the use of almost any commonly used measure of inequality. The choice of measure will naturally introduce additional implications and properties. Given this flexibility, an analyst can choose which measure to utilize in order to satisfy the additional properties he or she thinks important. Thus, for example a researcher who wishes to treat sequence inequality as being decreased more in a situation where an exchange of ranks happens between members of different groups, each of whom has lower ranks to begin with, can choose an inequality measure which shows the required form of transfer sensitivity (e.g., a generalized entropy index with appropriately chosen parameters).

First, we provide a very simple illustration of the application of our measure (assumed to be normalized so that its maximal value is one) to the synthetic examples described in the graphs above and then to real data. Table 1 suggests what values for each measure of group inequality would be realized in the situation described in each graph.

Table 1. *Component measures and polarization levels for selected graphs*

Figure	*RI*	*SI*	*GIC*	*P*
1	0	0	0	0
3	1	∼0	∼0	∼0
5	1	→1	>0	>0
11	1	→1	1	→1

In each case, polarization rises as each of the components become larger. In Figure 1, all components show zero group based inequality (and thus polarization is also zero), while in Figure 11, all measures of inequality approach their maximum levels (and thus polarization approaches its maximum value).

There is great interest in the differences in economic and social outcomes between identity groups and in the possible effects of such differences. In particular, researchers have focused on the propensity for inter-group differences to create a potential for violent civil conflict and for inter-state conflict. [15] Østby (2008) has attempted to empirically assess these relationships across countries using Demographic and Health Surveys (DHS) data, calculating measures of inter-group differences for diverse countries and attempting to relate them to measures of conflict. In order to establish a basis for comparison, we choose our second illustrative example accordingly, and estimate the level of representational inequality, sequence inequality, and the index of group inequality comparison for identity groups in five countries (Burkina Faso, Ghana. India, Namibia, and Peru) surveyed by the DHS in various years (1992 or 1993 and 1998 or 2000). Using this data allows us to make a valuable comparison between our measures and various others that are available in the data appendix provided by Østby (2008). One purpose of this exercise is to demonstrate the value of each of these concepts in understanding the differing nature of group based inequalities in different societies. We will show that each of the measures of inter-group inequality we calculate leads to different conclusions concerning the ordinal ranking and relative magnitude of inter-group inequalities in different countries and years. This is true of the four measures we have introduced here (*RI*, *SI*, *GIC*, and *P*) as well as of the measures originally calculated by Ostby. There is an especially stark difference between the "non-collapsed" measures which explicitly take

Figure 11. *Maximal polarization.*

account of identity groups and the "collapsed" measures which do not.

We estimate our measures through the use of a Stata module developed for the purpose. [16] Following Østby (2008) the attribute of interest is taken to be the asset level of the household in which the individual lives, and the identity group of interest is the "ethnic" category to which he or she belongs, although this should be more accurately understood as an ethnic, linguistic, or caste based grouping. Appendix A provides a list of groups in each country–year and the number of individual households in each country–year. The household asset level index is calculated on the basis of the following variables from the DHS surveys: *v119–v125* (dummy variables for whether or not each household has electricity, a radio, a television, a refrigerator, a bicycle, a motorcycle, and/or a car), and is calculated as a simple average of these.

Representational inequality in this context concerns the degree to which individuals from a given group are "localized" in the asset spectrum. The "locality" in this example is taken to be the quintile of the asset distribution. [17]

Table 2 provides the values of *RI, SI*, and *GIC* for our data. As mentioned, *RI* is calculated for each group and a weighted average of these generates the overall index of *RI*. The measure of inequality used when calculating all three underlying inter-group inequality measures is the *Gini* coefficient. Our measure of polarization (*PRJ*) is calculated as $[(RI)(SI)(GIC)]^{1/3}$. In the last two columns we provide, for purposes of comparison, measures of Polarization calculated by Østby (2008) for the same data. The first measure (*PO*1) refers to the influential measure proposed by Esteban and Ray (1994) and used by Østby as a measure of "economic polarization." This is a measure which does not take account of identities directly but does so indirectly by understanding the identity groups as given by location in the asset space. The second measure (*PO*2) refers to a modified form of the polarization measure designed to capture what Østby terms "ethnic/economic polarization" in which groups are explicitly taken into account. These measures share with ours the property of being a unitary measure generating a single number. The measure we present, however, has the additional advantage of being devised from separately informative constituent building blocks. These constituent blocks provide information that is missing or hidden by the modified measure proposed by Østby, as we shall show.

Turning now to the data, it is evident that for all measures there is significant variation in the observed levels of group based inequality. However there is a near perfect rank ordering of countries in their levels of polarization in the sense that the ranking of countries is largely independent of the years chosen. Namibia has the largest value while Ghana has the smallest for *PRJ* as well as *PO*2, the two non-collapsed measures of polarization. The level of polarization in Namibia is over twice as high as that in the other African countries with a similar demographic structure, Ghana and Burkina Faso. This finding clarifies that there need not be a simple relationship between the number and size of groups in the population and the level of group based inequality that is observed. Measures of group based difference such as ethnic fractionalization would, by contrast, suggest that these countries were very similar.

Why is Namibia the most polarized country among all countries? It is evident upon inspection that in 1992 this is primarily because *GIC* is the largest in the grouping. This in turn suggests that Namibia's high level of polarization does not derive primarily from inter-country differences in the degree of group segregation (reflected in *RI*) or group clustering (reflected in *SI*). Rather, it is the differences in the measure of group inequality contribution which cause Namibia to have higher levels of polarization. Analysis of this kind is not possible with the unitary measure proposed by Østby but is feasible when employing the composite measure we have introduced.

Peru has by far the highest level of representational inequality in the sample. This in turn suggests that groups are far more segregated by asset quintile in Peru than elsewhere. However, these groups are not as "far apart" as in Namibia, thereby making overall polarization lower. This hypothesis is further strengthened when one notes that the *Gini* coefficients as well as *GIC* for Peru are lower than anywhere else, suggesting that both individuals and groups are relatively not very far apart in their possession of assets. India similarly displays medium to high segregation (*RI*) and clustering (*SI*). However, inter-group inequality accounts for a larger share of total inequality than in Peru, contributing to higher levels of polarization.

These examples show the relevance of the fact that the constituent measures do not always move in step and therefore that a composite index might provide greater insights into the patterns of between group inequality than might be available if one were to use only one measure. The same is true of analyzing differences within countries over time. In Namibia in 2000, for example, while there was greater inequality between groups as measured by the level of representational inequality and sequence inequality than in 1992, there was a lower level of the index of group inequality comparison. This suggests that while there was a greater level of segregation and clustering of groups along the asset scale, the degree to which the

Table 2. *Segregation, clustering, and polarization in five countries*

Country	Year	*RI*	*SI*	*GIC*	*PRJ*	*PO*1	*PO*2	*Gini*
Namibia	1992	0.250	0.091	0.403	0.210	0.086	0.037	0.358
Namibia	2000	0.266	0.121	0.285	0.209	0.072	0.035	0.281
India	1998	0.212	0.101	0.295	0.185	0.064	0.035	0.635
Peru	1992	0.572	0.051	0.141	0.160	0.047	0.022	0.313
India	1992	0.273	0.075	0.182	0.155	0.076	0.033	0.680
Peru	2000	0.544	0.074	0.085	0.151	0.049	0.032	0.296
Burkina Faso	1998	0.187	0.064	0.215	0.137	0.066	0.016	0.864
Burkina Faso	1992	0.143	0.052	0.134	0.100	0.066	0.012	0.808
Ghana	1998	0.138	0.038	0.164	0.095	0.094	0.008	0.508
Ghana	1993	0.131	0.043	0.147	0.094	0.123	0.009	0.479
Correlation with PRJ						−0.38	0.92	−0.44

Key: *RI*, representational inequality; SI, sequence inequality; *GIC*, group inequality contribution; PRJ, polarization (Reddy and Jayadev (2009)), *PO*1, economic polarization (Østby, 2008); *PO*2, ethnic/economic polarization (Østby, 2008); *Gini*, Gini (Østby, 2008).

clusters were distant from each other fell. Such a hypothesis is further strengthened by looking at the *Gini* coefficient, which fell between the two years, suggesting that the distribution as a whole narrowed, as well as by examining *GIC*, which also fell significantly. The counteracting forces between the constituent elements appear to have nearly canceled each other out, as the observed level of polarization is almost exactly the same in the two years.

These examples serve to demonstrate the usefulness of the concept of polarization and in particular the importance of the composite construction to the researcher or policy-maker. Consider, for example, the analyst who observes the levels of inequality in Burkina Faso and in Peru. The fact that the *Gini* coefficient is *much* higher in the former than in the latter might initially suggest that social cohesion might be more tenuous in Burkina Faso than in Peru. However, an examination of the polarization index suggests that group based inequalities may be more salient in the latter than in the former. Furthermore, the fact that *RI* is much higher in Peru suggests that groups are much more segregated in their asset ownership (even though these are less unequally held overall). As such, social policy aimed at reducing inter-group differences in Peru may have more importance in maintaining civic peace and reducing inter-group conflict.

One might have speculated that there would be a broad correspondence between polarization as defined in our paper and polarization as assessed in Østby (2008). The result depends on the measure used in that paper. When the polarization measure employed is one that is "collapsed" (see Reddy & Jayadev, 2009) in the sense that the groups used in the analysis are defined by attribute levels themselves, the correlation is negative (−0.37). When the polarization measure employed is "non-collapsed" in the sense that the groups used in the analysis are defined by identity group, and thus closer to our understanding of polarization, the correlation is very high (0.92) although there are still notable discrepancies.

These results suggest that empirical researchers seeking to utilize measures of polarization need to be at least aware that there are multiple meanings of polarization in the literature as well as distinct ways of measuring these which have potentially enormous implications. It is interesting to note, for instance, that the time trends in polarization identified for the five countries in our sample when the measure we propose appear unrelated to the direction to the time trends in polarization identified by Østby's "economic polarization" (collapsed) measure for every one of the countries. The discrepancy in results is not surprising in light of the fact that the measure used does not explicitly take account of the partition of the society into identity groups but rather understands the groups being compared as being defined by the attribute itself. This is a common method of measuring polarization employed in the applied literature but has limited appeal for those seeking to understand the degree to which inter-identity-group differences are salient in interpreting social dynamics.

4. INTERPRETATION AND JUDGMENT IN THE APPLICATION OF THE MEASURES

Any empirical application of the concepts introduced above requires by its very nature the partitioning of the attribute space in some way so that it can be seen when members of distinct groups possess different levels of an attribute. For example, representational inequality concerns the extent to which particular attributes (whether income levels, occupations, or locations of residence) are shared by members of different groups. It is evident that this determination will depend on how these attributes are defined. For example, in an analysis of residential racial segregation in a city, defining the neighborhood of residence in the broadest way (to encompass the entire city) will lead to the conclusion that there is no racial segregation at all, since all races are represented in the same way that they are represented in the city as a whole. At the opposite extreme, defining the neighborhood of residence to be the individual household may lead to the conclusion that there is almost complete racial segregation if individuals in households are overwhelmingly from a single race. The appropriate way to define the neighborhood will lie between these extremes and will depend on the form in which data are available as well as the interests and purposes of the researcher. The fact that judgments as to the appropriate "bin size" are needed in empirical work is not therefore a deficiency but rather is intrinsic to the exercise, regardless of the measures used.[18] In our example in Section 4 we chose a particular bin (each quintile), but there are several other equally valid formulations.

Judgment on the part of investigator is, in fact, required in all parts of such an empirical exercise, especially in the very basic step of determining the appropriate partitioning of society into identity groups. The exercise we undertake in the previous section, for example, assesses group based inequalities based on a particular classification of a given society into groups. There are of course multiple ways in which the classification could have been done alternatively—instead of the broad designations of ethnicity used (e.g., caste in India) one could have used others (e.g., religious designations in India) for example—which will result in different values of the measures. This is a corollary to the recognition that individuals possess multiple identities simultaneously (on which see Sen, 2007): the choice of which identity group to assign may not be "natural." The point is elementary and should not need stating, but often it appears that applied work (including that published in the most august journals) ignores the issue.

Even if the choice of partition may appear uncontroversial, the features of the society which are of interest (e.g., inter-ethnic antagonisms or propensity to ethnic conflict) may need to be assessed with care. As an example, let us assume that there exists a society in which one could easily tell "blacks" from "whites." It would not necessarily follow that the interests of the members of these groups were opposed. Some "blacks" might have interests closely tied to those of some "whites," or indeed this distinction may have no social salience at all, frustrating such an attempt to interpret the society's dynamics.[19]

The mechanical interpretation of measures of the kind we introduce above could lead to mistaken conclusions concerning the features of the distribution which are being studied. While our measures are silent on the question of the appropriateness of a specific partition, it can accommodate any particular partition in its analysis (e.g., one could calculate these measure for "whites," "blacks" and "mixed race" as opposed to "whites" and "non-whites"). The choice of the appropriate partition will depend on the analyst's purpose.

A measure is meant to capture the extent of something. If the measure is to be compared across cases (whether these are defined geographically, temporally, counterfactually, or in some other way) that which is being measured must be deemed to be identical when the measure is the same. However, whether the measure possesses the same meaning in distinct contexts will depend on whether the features of those contexts are similar or different in a relevant sense, and this will in turn depend on the kind of meaning being emphasized.

For example, a one kilogram bar of gold would contain the same number of molecules on earth as it would on Mars. However, the speed with which such a bar would hit the ground when pushed from a 10 meter ledge would be different on Earth than it would on Mars (since the force of gravity is different on the two planets). Similarly, as mentioned above, whether two societies that have the same measure of polarization should be deemed to be socially divided to the same extent will depend on whether the social salience attached to group membership is the same in both cases. Comparisons across societies and across time must necessarily confront this issue. The current fashion of measuring polarization, segregation, and other such features appears often to elide these considerations in attempting to achieve a (false) rigor.

In using in our examples above a multiplicative measure for polarization, we are employing just one way in which information about underlying features of the distribution can be aggregated. It should be perhaps emphasized that each constituent measure provides distinct information and as such, examining each measure is valuable in assessing the underlying salience of identity in the distribution of the attribute. Representational inequality provides a metric of the isolation of identity groups in pre-specified achievement brackets, sequence inequality provides a metric of the degree to which identity groups are hierarchically ordered, and group inequality comparison provides a metric of the degree to which the membership of distinct groups accounts for overall inequality. This distinctness means that excluding any one of these or solely employing an aggregative measure could lead to the loss of important insights.

As an example, consider Figure 2 depicting maximal representational inequality. Such a situation is perfectly consistent with a polarization value of zero. To see this imagine that the mean income of blacks and whites is the same, which makes the value of group inequality comparison equal zero. It is easy to construct other such situations in which some of the constituent measures are non-zero and even large but polarization is zero. These examples show the importance of taking note of each kind of inter-group inequality in forming judgments about the extent of social divisions.

5. CONCLUSION

Social situations can differ in the extent to which members of a group share experiences with members of other groups (representational inequality), experience the same or different relative positions (sequence inequality), and experience differences in the extent to which interpersonal inequalities are accounted for by inter-group differences (group inequality comparison). These concepts are distinct but complexly interrelated. They each integrate empirical observations and evaluative judgments, and can in turn be integrated to form a measure of the overall polarization of societies through the further application of such judgments.

Social scientists undertaking empirical work on social divisions and their relevance to understanding social phenomena such as ethnic conflicts cannot be oblivious to the choice between alternate concepts and measures which is present in this field. That choice is replete with implications for the assessment of the relative extent of inter-group differences in different societies. The measures we propose have simple interpretations and can be used to distinguish between various underlying forms of social division.

Although there is indeed, as noted by Amartya Sen above, "a long way to go still to make adequate social sense of economic measures," the approach we present can play a part in this endeavor.

NOTES

1. It should be noted that some of these concepts, especially Polarization, have several different meanings and do not always refer to the same underlying idea. For example, in Wolfson (1994), Polarization is the absence of "middleness" in an income spectrum, in Esteban and Ray (1994) polarization refers the clustering of individuals at various points of an income space, in Montalvo and Reynal-Querol (2005), polarization refers to the degree to which a distribution of ethnic groups is "distant" from a situation in which the population is equally split between two ethnic groups, and in Østby (2008), polarization is related to differences in the average achievement of groups "weighted by the degree of inward identification of each group."

2. This is done in the interest of space. A reader wishing to engage with a more formal presentation of these ideas is referred to Reddy and Jayadev (2009).

3. Segregation is defined by the Oxford English Dictionary, *inter alia,* as "The separation of a portion of portions of a collective or complex unity from the rest; the isolation of particular constituents of a compound or mixture."

4. A cluster is defined by the Oxford English Dictionary as, *inter alia*, "A collection of things of the same kind…growing closely together; a bunch… a number of persons, animals, or things gathered or situated close together; an assemblage, group, swarm, crowd."

5. For a discussion on another way of understanding the *GIC* concept, see Reddy and Jayadev (2009).

6. The Oxford English Dictionary defines the verb "polarize" as "To accentuate a division within (a group, system, *etc.*); to separate into two (or occas. several) opposing groups, extremes of opinion, *etc.*"

7. See Jayaraj and Subramanian (2006) for an example of such an approach.

8. In the case of multiple groups, the transfer principle is defined in terms of a progressive (regressive) "balanced bilateral transfer principle" in which two matched transfers take place. This involves the movement of a population mass of individuals from a group which is over (under) represented at an income bracket to one in which they are under (over) represented combined with a corresponding shift of an equal population mass of individuals from another group which are over (under) represented at the second income bracket to the first income bracket in which they are under (over) represented. It can be shown that in the absence of equiproportional representation, such a transfer can always take place (see Reddy & Jayadev, 2009).

9. This may be readily seen by noting that the white with the highest income dominates 25 blacks, that the white with the next highest income dominates 24 blacks and so on. The number 325 results from summing the consequent series according to the formula $((n)(n+1)/2)$.

10. It can be shown that an exact and monotonic relationship exists between the concepts of group rank dominance and of average rank. Both of these could be seen to be indicators of the placement of groups in the

attribute hierarchy (in the extreme complete clustering of groups) and will thus be referred to as indicators of a group's rank sequence position. See Reddy and Jayadev (2009) (in particular, Appendix 2).

11. Indeed, ways of viewing group differences that are not based on representative incomes alone can also be envisioned; for example methods which include the examination of the extent of "non-overlap" between distributions or comparisons of higher moments of the group-specific distributions of incomes. For a wide-ranging discussion of methods of defining group separation, see Anderson (2004, 2005).

12. These can be partial orderings based on dominance rankings of the vectors (2-tuples or 3-tuples) defined by the individual measures of inter-group differences or can be complete orderings if based on some method of aggregation of these measures.

13. For an axiomatic description of this measure see Reddy and Jayadev (2009).

14. In particular, the sum of "between-group" and "within group" inequalities is equal to the overall interpersonal inequality in the society for such measures. Yitzhaki (1994) shows that the Gini coefficient is additively decomposable in a different sense: it is the sum of a "between," a "within" and an "overlapping" component.

15. Indeed, the Journal of Peace Research recently published a "**Special Issue on Polarization and Conflict**" (Journal of Peace Research, 2008, Vol. 45, No. 2, edited by G. Schneider and J. Esteban) **which has addressed several of these themes**.

16. The code is available from the authors upon request. We thank Rahul Lahoti for creating it.

17. Of course, the description of the bins into which attributes are placed will have an impact on the measured level of representational inequality for the group: the level of representational inequality among groups when each bin is defined as quintiles will not in general be the same as the level of representational inequality based on another categorization.

18. The selection of the appropriate bin size (or more generally the boundaries of classification) must be itself based on appropriate principles, which it may be necessary appropriately to weight or to select among. For instance, one might choose bin size to correspond to the income spanned by a given quantile of the population, a fixed income interval, the income interval corresponding to a given proportion of the span of incomes for the entire populations, or in still other ways (for a recent interesting example see Fryer and Echenique (2007)).

19. For an apposite example see Malcolm X, "The House Negro and the Field Negro," Speech to Workers, Selma, Alabama, Feb. 4th, 1965.

REFERENCES

Anderson, G. (2004). Toward an empirical analysis of polarization. *Journal of Econometrics, 122*(2004), 1–26.

Anderson, G. (2005). *Polarization*. Working paper, University of Toronto. <http://www.chass.utoronto.ca/≃anderson/Polarization.pdf>.

Duclos, J-Y., Esteban, J., & Ray, D. (2004). Polarization: Concepts, measurement, estimation. *Econometrica*, 1737–1772.

Esteban, J.-M., & Ray, D. (1994). On the measurement of polarization. *Econometrica, 62*, 819–851.

Fryer, R., & Echenique, F. (2007). A measure of segregation based on social interactions. *Quarterly Journal of Economics, 122*(2), 441–485.

Jayaraj, D., & Subramanian, S. (2006). Horizontal and vertical inequality: Some interconnections and indicators. *Social Indicators Research, 75*(1), 123–139.

Journal of Peace Research (2008), Vol. 45, No. 2.

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict and civil war. *American Economic Review, 95*(3), 796–816.

Østby, G. (2008). Polarization, horizontal inequalities and violent civil conflict. *Journal of Peace Research, 45*(2), 143–162.

Reddy, S. G., & Jayadev, A. (2009). *Inequalities and identities*. Unpublished working paper available on: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1162275>.

Sen, A. (2006). Conceptualizing and measuring poverty. In D. B. Grusky, & S. M. Ravi Kanbur (Eds.), *Poverty and inequality*. Stanford, CA: Stanford University Press.

Sen, A. (2007). *Identity and violence: The illusion of destiny*. New York/London: WW Norton & Company.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica, 48*(3), 613–625.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica, 50*, 3–17.

Stewart, F. (2001). *Horizontal inequalities: A neglected dimension of development*. WIDER Annual Lectures 5, UNU WIDER.

Subramanian, S. (2009). *Alternative approaches to reckoning group disparities in distributional analysis: Concepts, measurement, and illustrative applications*. Madras Institute of Development Studies, Mimeo.

Wolfson, M. C. (1994). When inequalities diverge. *American Economic Review Papers and Proceedings, 94*, 353–358.

X., M (1965), "The House Negro and the Field Negro", Speech to Workers, Selma, Alabama, Feb. 4th, 1965. available at: http://www.iowalakes.edu/Directories/faculty/burns/informative/Malcolm_X%27s_House-Field_Negro_speech.htm.

Yitzhaki, S. (1994). Economic distance and overlapping distributions. *Journal of Econometrics, 61*(1), 147–159.

Zhang, X., & Kanbur, R. (2001). What difference do polarisation measures make? An application to China. *Journal of Development Studies, 37*, 85–98.

APPENDIX A

Ethnicity (linguistic or caste grouping) 1992 or 1993	Sample size in 1992 or 1993	Ethnicity (linguistic or caste grouping) 1998 or 2000	Sample size in 1998 or 2000
Burkina Faso			
Bobo	499	Bobo	959
Dioula	483	Dioula	128
Fulfude (Peul)	288	Fulfulde/Peul	1,178
Gourmantche	215	Gourmatche	1,423
Gouroussi	363	Gourounsi	959
Lobi	264	Lobi	302
Mossi	3,649	Mossi	11,644
Senoufo	83	Senoufo	529
Other	326	Bissa	815

Appendix A—Continued

Ethnicity (linguistic or caste grouping) 1992 or 1993	Sample size in 1992 or 1993	Ethnicity (linguistic or caste grouping) 1998 or 2000	Sample size in 1998 or 2000
Touareg Bella	69	Touareg/Bella	90
Dk	2	Dafing	324
Total	6,241	Dagara	197
		Samo	451
		Other Burkina	662
		Nigerian	16
		Togolaise	20
		Beninoise	4
		Malian	34
		Ivory Coast	8
		Ghanian	11
		Other African	11
		Total	19,765
Ghana			
Asante	741	Asante	671
Akwapim	147	Akwapim	172
Fanti	553	Fante	574
Other Akan	827	Other Akan	823
Ga.Adangbe	364	Ga/Adangbe	344
Ewe	679	Ewe	646
Guan	99	Guan	71
Mole-Dagbani	710	Mole-Dagbani	510
Grussi	161	Grussi	202
Gruma	106	Gruma	374
Hausa	31	Hausa	66
Other	88	Dagarti	288
Total	4,506	Other	102
		Total	4,843
India			
Scheduled caste	45,679	Scheduled caste	64,352
Scheduled tribe	55,406	Scheduled tribe	52,449
Other	303,966	Other backward caste	109,000
		None of them	167,243
Total	405,051	Total	393,044
Namibia			
English	33	Afrikaans	895
Afrikaans	516	Damara/Nama	1,473
Oshivambo	2,451	English	52
Damara/Nama	657	Herero	854
Herero	280	Kavango languages	540
Kwangali	228	Caprivi languages	289
Lozi	307	Oshiwambo	2,446
Tswana	21	San	108
San	46	Tswana	25
German	23	Other	72
Other	852	Total	6,754
Total	5,414		
Peru			
Castellano	14,558	Spanish	82,006
Quechua	1,097	Quechua	10,000
Aymara	116	Aymara	1,009
Other local language	63	Others	677
Foreign language	1	Abroad Idioma	45
Total	15,835	Total	93,737