

IOWA STATE UNIVERSITY

Measuring Segregation

David M. Frankel, Oscar Volij

May 2007

Working Paper # 07009

Department of Economics Working Papers Series

Ames, Iowa 50011

Iowa State University does not discriminate on the basis of race, color, age, religion, national origin, sexual orientation, gender identity, sex, marital status, disability, or status as a U.S. veteran. Inquiries can be directed to the Director of Equal Opportunity and Diversity, 3680 Beardshear Hall, (515) 294-7612.

Measuring Segregation*

David M. Frankel

Oscar Volij

Iowa State University

Iowa State University

and Ben Gurion University

March 19, 2007

Abstract

We propose a set of axioms for the measurement of school-based segregation with any number of ethnic groups. These axioms are motivated by two criteria. The first is evenness: how much do ethnic groups' distributions across schools differ? The second is representativeness: how different are schools' ethnic distributions from one another? We prove that a unique ordering satisfies our axioms. It is represented by an index that was originally proposed by Henri Theil (1971). This "Mutual Information Index" is related to Theil's better known Entropy Index, which violates two of our axioms.

Journal of Economic Literature Classification Numbers: C43, C81, D63.

Keywords: Segregation, indices, measurement, peer effects, schools, education, equal opportunity.

*Email addresses: dfrankel@econ.iastate.edu; volij@iastate.edu. We thank seminar participants at University of the Basque Country, Brown University, University of Haifa, University of Iowa, Iowa State University, Rice University, Tel Aviv University, the 2004 Summer Meetings of the Econometric Society, the 2007 Annual Meetings of the American Real Estate and Urban Economics Association, and the Murat Sertel Memorial Conference on Economic Theory. Volij thanks the Spanish Ministerio de Educación y Ciencia (project SEJ2006-05455) for research support.

1 Introduction

Segregation is a pervasive social issue. The segregation of men and women into different occupations helps explain the gender gap in earnings.¹ Racial segregation in schools is thought to contribute to low educational achievement among minorities.² Residential segregation has been blamed for black poverty, high black mortality, and increases in prejudice among whites.³ In other contexts, segregation is viewed more positively. The formation of homogeneous living areas has been discussed as a solution to highly polarized conflicts in the Middle East, Yugoslavia, and elsewhere.

The literature on segregation measurement has generated over 20 different indices (see Massey and Denton [37] and Flückiger and Silber [20]). While some papers have analyzed the properties of various indices, very few of them have provided a full characterization, and none of these have used purely ordinal axioms. Further, the existing characterizations treat only the two-group case. In this paper we provide a full ordinal characterization of a segregation index for the multigroup case.

Axiomatizations are important because they characterize an index in terms of basic properties and thus facilitate the comparison of different measures. Ordinal axioms are more appealing than cardinal ones because they refer to bilateral comparisons and not to their specific functional representations. Multigroup segregation orderings are important because they allow us to study units (cities, school districts, etc.) with more than two ethnic groups and to compare units with different numbers of groups.

In this paper we focus on contexts in which geography is unimportant. In some cases, such as residential neighborhoods, this might be a strong assumption. In others, it is more innocuous. For instance, the presence of other schools near a given student's school typically does not have a great effect on the student's educational outcomes. Hence, our presentation will focus on school district segregation.

¹See Cotter et al [14], Lewis [34], and Macpherson and Hirsh [36].

²Recent studies include Boozer, Krueger, and Wolkon [3] and Hanushek, Kain, and Rivkin [25].

³See Cutler and Glaeser [16], Collins and Williams [12], and Kinder and Mendelberg [33], respectively.

Formally, we define a segregation ordering as an ordering on school districts: a ranking from most segregated to least segregated. We propose a set of axioms that, we argue, such an ordering should satisfy. We then prove that there is a unique ordering that satisfies our axioms. It is represented by a simple index: the total entropy of the school district, minus the within-school entropy. We call this the “Mutual Information” index. It can be interpreted as the average amount of information a student’s school reveals about her ethnicity.

The Mutual Information index was first proposed by Theil [52, p. 653] and was applied by Fuchs [23] and Mora and Ruiz-Castillo [38, 41] to study gender segregation in the labor force.⁴ It is related to the more widely used Entropy index (Theil [53]; Theil and Finizza [54]), which equals one minus the ratio of within-school entropy to total entropy. While the Entropy index is normalized to reach a maximum value of one, the Mutual Information index has no maximum value. However, the Entropy index violates two of our axioms.

In order to judge our axioms, one must have an idea of what we are trying to measure. A starting point is James and Taeuber’s [32] definition of segregation as the tendency of ethnic groups to have different distributions across locational units such as schools or neighborhoods. In a later paper, Massey and Denton [37] discern five different dimensions of segregation. The first, evenness, agrees with James and Taeuber’s definition. The second dimension is isolation from the majority group. The three other dimensions rely on geographic information and thus are not relevant to our study.⁵

While evenness generalizes easily to the multigroup setting, isolation is more of a challenge, since there is more than one other ethnic group from which a student can be “isolated.” Hence, we replace isolation with the related concept of representativeness: to what extent do students attend schools that have different ethnic compositions than the district as a whole? The concepts are related, since racially isolated schools are, by definition, not representative

⁴See also Herranz, Mora, and Ruiz-Castillo [26]. Some of the properties of the Mutual Information index have been previously noted by Mora and Ruiz-Castillo in the case of two ethnic groups [39, 40].

⁵These dimensions are concentration in a small area, centralization in the urban core, and clustering in a contiguous enclave.

of their districts. But unlike isolation, representativeness is not based on exposure to just one other group.

The concept of representativeness is connected to economic issues such as equality of opportunity. Boozer, Krueger, and Wolkin [3] and Hoxby [27] find that the ethnic composition of a school affects individual students' achievement. In the presence of such ethnic-based peer effects, a lack of representativeness can create unequal educational opportunities among students of different races. Evidence for this appears in Hanushek, Kain, and Rivkin [25], who find that the higher proportion of blacks in the school attended by the typical black student can explain a large portion of the black-white wage gap.

Representativeness and evenness are dual concepts. In Table 1 we depict a school district as a matrix, where the rows are ethnic groups, the columns are schools, and the cells contain numbers of students. A deviation from evenness (representativeness) corresponds to differences in the row (column) percentages. The Mutual Information index treats these deviations symmetrically: if the matrix is transposed, the Mutual Information of the district is unchanged. This property is a result of our axiomatization rather than an assumption.

| District X | | |
|------------|----------|----------|
| | School A | School B |
| Blacks | 500 | 200 |
| Whites | 100 | 400 |

Table 1: Matrix representation of a district.

We study segregation with respect to a particular choice of locations (schools, classrooms, etc.). The segregation ranking of a district is sensitive to this choice. However, one can study segregation at several nested levels at once by exploiting the additive decomposability of the Mutual Information index: segregation between the classrooms of a district equals segregation between the district's schools plus the population-weighted mean level of segregation within the schools (Mora and Ruiz-Castillo [39]). This property is not satisfied by the other common segregation indices.

We also study segregation with respect to a particular ethnic schema. Sensitivity to

the choice of ethnic schema is a property of any nontrivial segregation measure.⁶ However, the Mutual Information index is also decomposable with respect to ethnic groups. For instance, if students are classified by both race and language spoken at home, then total segregation equals segregation by race plus the population-weighted mean level of segregation by language within the racial groups. This property can be used to study segregation with several nested ethnic schemas simultaneously and is not satisfied by other common indices.

In our analysis, we will assume that each ethnic group can be distributed in arbitrary real proportions across the schools in a district. This assumption ensures, by appropriately distributing students, that all schools can be representative of the district, i.e., be small copies of it. It is a good approximation when the ethnic groups have many more members than there are schools in the district and when capacity constraints on schools are not binding. It would not be suitable, for instance, if there were three equal-size ethnic groups and a maximum capacity of two students per school; or if there were an ethnic group with two members to be allocated among three schools. These are not the intended applications of our model.

The rest of the paper is organized as follows. Notation is introduced in section 2. In section 3 we explain our axioms. The main result appears in section 4. In section 5, we survey other multigroup segregation indices and consider three other properties that an index might satisfy. We survey related literature in section 6. Proofs are collected in an appendix.

2 Notation

Formally, we define a (school) district as follows:

Definition 1 *A district X consists of*

⁶Consider a district with two schools, one with 50 Hispanic whites and the other with 50 Anglo whites. Ignoring Hispanic origin, the district is completely integrated; taking it into account, the district is completely segregated.

- A nonempty and finite set of ethnic groups $\mathbf{G}(X)$
- A nonempty and finite set of schools $\mathbf{N}(X)$
- For each ethnic group $g \in \mathbf{G}(X)$ and for each school $n \in \mathbf{N}(X)$, a nonnegative number T_g^n : the number of members of ethnic group g that attend school n .

For instance, in the district X depicted in Table 1, $\mathbf{G}(X) = \{Black, White\}$, $\mathbf{N}(X) = \{A, B\}$, $T_{Black}^A = 500$, and so on. The district in Table 1 is depicted in matrix format. We will also sometimes specify a district in *list format*: $\langle (T_g^n)_{g \in \mathbf{G}} \rangle_{n \in \mathbf{N}}$. For instance, $\langle (1, 2), (3, 1) \rangle$ denotes a district with two ethnic groups (e.g., blacks and whites) and two schools. The first school, $(1, 2)$, contains one black and two whites; the second, $(3, 1)$, contains three blacks and one white. For any two districts X and Y , $X \uplus Y$ denotes the result of combining the schools in X and the schools in Y into a single district.⁷ If X is a district and α is a nonnegative scalar, then αX denotes the district in which the number of students in each group and school has been multiplied by α ; for instance, if $X = \langle (1, 2), (3, 1) \rangle$, then $2X = \langle (2, 4), (6, 2) \rangle$. Also, $c(X)$ denotes the district that results from combining the schools in X into a single school.

⁷Formally, $X \uplus Y$ denotes the district $\langle (T_g^n)_{g \in \mathbf{G}} \rangle_{n \in \mathbf{N}}$, where $\mathbf{G} = \mathbf{G}(X) \cup \mathbf{G}(Y)$ is the set of ethnic groups that are present in either district and $\mathbf{N} = \mathbf{N}(X) \cup \mathbf{N}(Y)$ is the set of all schools in the two districts. Naturally, if an ethnic group g is present only in one district, then T_g^n , the number of members of group g in school n , equals zero for all schools n in the other district.

The following notation will be useful:

$$T_g = \sum_{n \in \mathbf{N}} T_g^n: \text{ the number of students in ethnic group } g \text{ in the district}$$

$$T^n = \sum_{g \in \mathbf{G}} T_g^n: \text{ the total number of students who attend school } n$$

$$T = \sum_{g \in \mathbf{G}} T_g: \text{ the total number of students in the district}$$

$$P_g = \frac{T_g}{T}: \text{ the proportion of students in the district who are in ethnic group } g$$

$$P^n = \frac{T^n}{T}: \text{ the proportion of students in the district who are in school } n$$

$$p_g^n = \frac{T_g^n}{T^n} \text{ (for } T^n > 0\text{): the proportion of students in school } n \text{ who are in ethnic group } g$$

The *group distribution of a district* X is the vector $(P_g)_{g \in \mathbf{G}}$ of proportions of the students in the district who are in each ethnic group. The *group distribution of a nonempty school* n is the vector $(p_g^n)_{g \in \mathbf{G}}$ of proportions of students in school n who are in each ethnic group. A school is *representative* if it has the same group distribution as the district that contains it.

3 Axioms

Let \mathcal{C} be the set of all districts. A *segregation ordering* \succcurlyeq is a complete and transitive binary relation on \mathcal{C} . We interpret $X \succcurlyeq Y$ to mean “district X is at least as segregated as district Y .” The relations \sim and \succ are derived from \succcurlyeq in the usual way.⁸ A related concept is the *segregation index*: a function $S : \mathcal{C} \rightarrow \mathbb{R}$. The index S *represents* the segregation ordering \succcurlyeq if, for any two districts $X, Y \in \mathcal{C}$,

$$X \succcurlyeq Y \iff S(X) \geq S(Y) \tag{1}$$

⁸That is $X \sim Y$ if both $X \succcurlyeq Y$ and $Y \succcurlyeq X$; $X \succ Y$ if $X \succcurlyeq Y$ but not $Y \succcurlyeq X$.

Every index S induces a segregation ordering that is defined by (1).

We impose axioms not on the segregation index but on the underlying segregation ordering. These approaches are not equivalent. As in utility theory, a segregation ordering may be represented by more than one index, and there are segregation orderings that are not captured by any index.

A district's *segregation ranking* or simply its *segregation* is its place in the segregation ordering. We will sometimes say that if a transformation $\sigma : \mathcal{C} \rightarrow \mathcal{C}$ is applied to a district X , then “the segregation of the district is unchanged” or “the district's segregation ranking is unaffected.” By this we mean that $\sigma(X) \sim X$. If this holds for all districts X , then we will say that the segregation in a district is invariant to the transformation σ .

Evenness and representativeness are properties of the row and column percentages of the district matrix. Nothing in these concepts suggests that the rows or columns should be treated asymmetrically. Accordingly, our first axiom states that the order of the schools or groups and their labels such as “black”, “Roosevelt School,” etc., do not matter: all that matters is the number of each group who attend each school.

Symmetry (SYM) The segregation in a district is invariant to any relabeling or reordering of the groups or the schools in the district.

One type of research for which this axiom may not be suitable is work that focuses on the problems that face a particular ethnic group. For instance, if one is interested in the social isolation of blacks from all other groups, then one may want to treat blacks differently (see, e.g., Echenique and Fryer [18]).

The criteria of evenness and representativeness pertain to the row and column percentages in the district matrix. Multiplying the whole matrix by a scalar does not affect these percentages, so it should not affect the segregation ranking of a district. Hence, we assume the following axiom.

Weak Scale Invariance (WSI) The segregation ranking of a district is unchanged if the numbers of agents in all ethnic groups in all schools are multiplied by the same positive scalar: for any district $X \in \mathcal{C}$ and any positive scalar α , $X \sim \alpha X$.

This axiom implies that the districts $X = \langle(10^6, 0), (0, 10^6)\rangle$ and $Y = \langle(100, 0), (0, 100)\rangle$ are equally segregated. One may argue that X is more segregated than Y because X is less likely to be the outcome of random assignment of students to schools (see, e.g., Cortese, Falk, and Cohen [13]). However, an important motivation for this paper is to produce a measure that can be used to study the effects of segregated schools on their students. In this context, realized segregation would appear to be the appropriate concept.⁹ For researchers who desire a test of random assignment, Mora and Ruiz-Castillo [40] show that a transformation of the Mutual Information index can be used for this purpose.

We motivate the next three axioms with a brief discussion of the concepts of within-cluster and between-cluster segregation. Suppose we partition a district into K clusters, C_1 through C_K , each consisting of a subset of schools in the district. Define segregation within a given cluster as the segregation ranking of the cluster viewed in isolation, as a distinct school district. Define segregation between the K clusters as the segregation ranking of a district with K schools $k = 1, \dots, K$, where school k consists of the students in cluster k in the original district. We would like the district's segregation ranking to be a function of segregation *within* each cluster, segregation *between* the clusters, and the relative sizes of the different clusters. Naturally, a district's segregation ranking should be a *nondecreasing* function of both segregation within each cluster and segregation between the K clusters.

The first axiom that uses this principle is illustrated in Figure 1. In panel *a*, we divide a district into two clusters. The first, cluster C_1 , consists of all schools except a single school n . The second, cluster C_2 , consists of school n alone. In panel *b*, school n has been torn down and replaced by two new schools, n_1 and n_2 . Each student who formerly attended school n now attends either school n_1 or n_2 ; all other students attend the same schools as before.

This change should not lower segregation in the district. Why? The only factor affected by the split is segregation within cluster C_2 . There has been no change in segregation within cluster C_1 , segregation between the clusters, or the relative sizes of the two clusters. Since

⁹For a colorful defense of realized segregation measures, see Taeuber and Taeuber [51, p. 886]

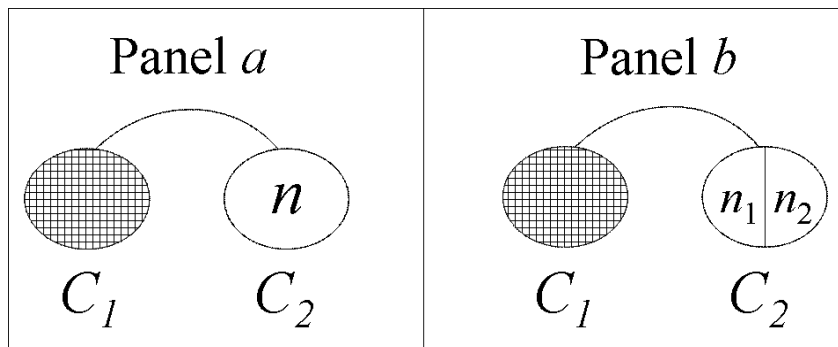


Figure 1: The School Division Property. In panel a , a district has been partitioned into two clusters, one containing a single school n . In panel b , school n has been divided into schools n_1 and n_2 . The School Division Property states that segregation is no lower in panel b than in panel a and, moreover, that segregation is the same in the two panels if schools n_1 and n_2 have the same ethnic distribution.

initially cluster C_2 was not segregated at all, splitting school n cannot lower segregation in this cluster. Accordingly, splitting school n should not lower segregation in the district either. If schools n_1 and n_2 have the same ethnic distribution, then cluster C_2 is not segregated at all after the split, since each school is representative of the cluster. In this case, the segregation ranking of the district should not change. These conclusions are formalized in the following axiom.

School Division Property (SDP) Let $X \in \mathcal{C}$ be a district in which the set of schools is \mathbf{N} . Let X' be the result of splitting some school $n \in \mathbf{N}$ into two schools, n_1 and n_2 . Then $X' \succcurlyeq X$. If both schools have the same ethnic distribution, then $X' \sim X$.

The School Division Property is related to two properties that are discussed by James and Taeuber [32] and subsequent authors. The first is *organizational equivalence*: if a school is divided into two schools that have the same group distribution, the district's level of segregation does not change. The second is the *transfer principle*. When there are two demographic groups, the transfer principle states that if a black (white) student moves from one school to another school in which the proportion of blacks (whites) is higher, then segregation in the district rises. In the case of two ethnic groups, SDP follows from

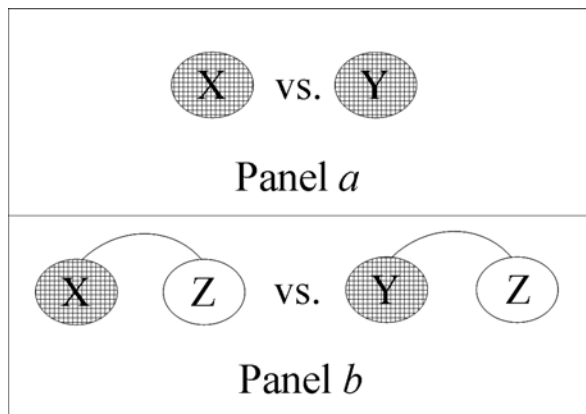


Figure 2: Type I Independence (IND1). Panel *a* shows two districts, X and Y , that have the same size and ethnic distribution. IND1 states that adjoining the same cluster containing a single school to the two districts (panel *b*) does not affect which district is more segregated.

organizational equivalence and the transfer principle.¹⁰ But while SDP applies directly with any number of groups, it is unclear what form the transfer principle should take with more than two groups.¹¹

Our next axiom is illustrated in Figure 2. In panel *a*, two districts, X and Y , are being compared. The districts are assumed to have the same number of students and ethnic distribution. In panel *b*, a cluster that contains a single school has been adjoined to each of these districts. The axiom states that the addition of this cluster should not affect which district is more segregated. That is, the district on the left hand side in panel *b* is more segregated than the district on the right hand side in panel *b* if and only if X is more segregated than Y .

Type I Independence (IND1) Let $X, Y \in \mathcal{C}$ be two districts with equal populations and equal group distributions. Then for any district Z that contains a single school, $X \succcurlyeq Y$ if and only if $X \uplus Z \succcurlyeq Y \uplus Z$.

¹⁰Proof available on request.

¹¹For instance, suppose a black student moves to a school that has higher proportions of both blacks and Asians but fewer whites. Since there are more blacks, one might argue (using the transfer principle) that segregation has gone up. On the other hand, blacks are now more integrated with Asians. One attempt to overcome this difficulty appears in Reardon and Firebaugh [45].

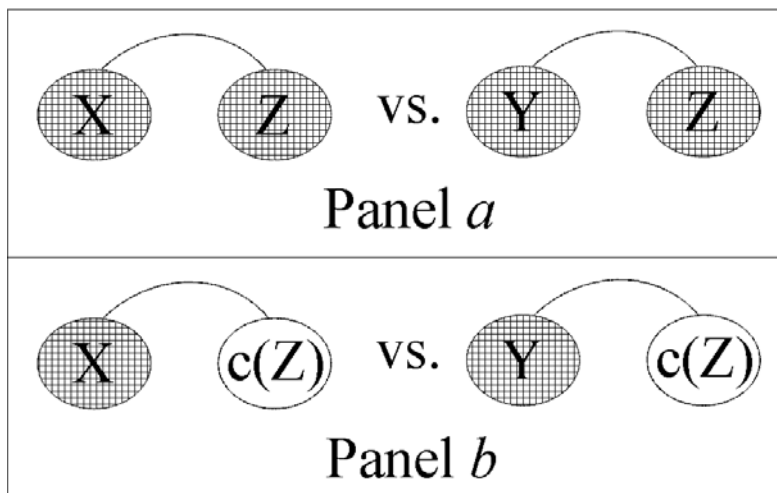


Figure 3: In panel a , a given district, Z , is combined with each of two districts, X and Y , which have the same total number of students but possibly different ethnic distributions. In panel b , all the schools in Z have been combined into a single school. Type II Independence states that this merger does not affect which combined district is more segregated.

For an intuition, we once again rely on the concepts of within-cluster and between-cluster segregation, where the clusters are now X , Y , and Z . Since X and Y have the same size and group distribution, in each combined district in panel b the *between-cluster* segregation is the same. Moreover, segregation *within* cluster Z is the same in the two combined districts. Hence, which of the combined districts in panel b is more segregated reduces to whether segregation within cluster X is greater than segregation within cluster Y .

A second type of independence is depicted in Figure 3. In panel a , a given district, Z , is paired with each of two districts, X and Y . As in Figure 2, X and Y have the same total number of students; unlike that case, their ethnic distributions may differ. In panel b , all the schools in Z have been combined into a single school; the resulting cluster is denoted $c(Z)$. Type II Independence states that this merger of schools does not affect which combined district is more segregated.

Type II Independence (IND2) Let $X, Y, Z \in \mathcal{C}$ be three districts such that $T(X) = T(Y)$. Let $c(Z)$ be the cluster that results from combining the schools in Z into a single school. Then $X \uplus Z \succcurlyeq Y \uplus Z$ if and only if $X \uplus c(Z) \succcurlyeq Y \uplus c(Z)$.

A motivation is as follows. Suppose that, in panel *a*, the combination of *X* and *Z* is more segregated than the combination of *Y* and *Z*. What must be driving this? Segregation *within* cluster *Z* is the same in the two districts in panel *a*. So the combination of within-*X* segregation and between-*X*-and-*Z* segregation must exceed the combination of within-*Y* segregation and between-*Y*-and-*Z* segregation. Moreover, since *X* and *Y* are of the same size, the relative importance of within-cluster and between-cluster segregation is the same in the two cases. Now consider panel *b*. Merging the schools in *Z* does not affect segregation between this cluster and either *X* or *Y*. Consequently, if in panel *b* the district containing cluster *X* is more segregated than the district containing cluster *Y*, then the combination of within-*X* segregation and between-*X*-and-*Z* segregation must exceed the combination of within-*Y* segregation and between-*Y*-and-*Z* segregation, just as in panel *a*. Moreover, since the merger does not affect the size of any cluster, it does not change the relative importance of within-cluster and between-cluster segregation. Accordingly, merging the schools in *Z* should *not* affect which merged district is more segregated. In other words, the degree of segregation within a given cluster should not affect the relative importance of between-cluster segregation and segregation within the other clusters in the district. In section 5 we show that if an ordering violates Type II Independence, then an index that represents it cannot be decomposable across schools in a particular simple way (Observation 1).

The next axiom is used to compare districts with different ethnic distributions. It states that segregation is invariant to the division of an existing ethnic group into two identically distributed subgroups. For instance, if white students are divided into those with blue eyes and those with brown, and these groups have the same distribution across schools, then the segregation of a district should not change.

Group Division Property (GDP) Let $X \in \mathcal{C}$ be a district in which the set of ethnic groups is G . Let X' be the result of partitioning some ethnic group $g \in G$ into two ethnic groups, g_1 and g_2 , such that both ethnic groups have the same distribution across schools: $\frac{T_{g_1}^n}{T_{g_1}} = \frac{T_{g_2}^n}{T_{g_2}}$ for all $n \in \mathbf{N}$.¹² Then $X' \sim X$.

¹²Note that X' has the same set N of schools as X and for each school $n \in N$, $T_g^n = T_{g_1}^n + T_{g_2}^n$.

A motivation is as follows. Suppose we partition the ethnic groups of X into K sets or “supergroups.” Define within-supergroup segregation to be the segregation of the district that would result if all students who are not members of the given supergroup were removed. Let between-supergroup segregation be the segregation of the district that would result from treating each supergroup as a single ethnic group. Then segregation in X should be a function of segregation *within* each supergroup, segregation *between* the supergroups, and the relative *sizes* of the supergroups.

This principle helps motivate GDP in the following way. Let us partition the ethnic groups of X into two supergroups, one consisting of group g alone and the other consisting of all other groups. Suppose group g is split into two groups, g_1 and g_2 , which have the same distribution across schools. This change clearly does not affect segregation within either supergroup, nor does it affect segregation between the supergroups or the relative sizes of the two supergroups. Hence, the district’s segregation ranking should not be affected by the split. In section 5 we show that an ordering that violates GDP cannot be represented by an index that is decomposable over groups in a particular way (Observation 1).

The next axiom is a technical continuity property. We rely on this axiom to prove that the segregation ordering is represented by a segregation index.

Continuity (C) Let $X, Y, Z \in \mathcal{C}$ be three districts. Then the sets

$$\{c \in [0, 1] : cX \uplus (1 - c)Y \succcurlyeq Z\} \quad \text{and} \quad \{c \in [0, 1] : Z \succcurlyeq cX \uplus (1 - c)Y\}$$
 are closed.

Our final axiom states that there exist two districts with two nonempty ethnic groups that are not equally segregated. It is needed to rule out the trivial segregation ordering.

Nontriviality (N) There exist districts $X, Y \in \mathcal{C}$, each with exactly 2 nonempty ethnic groups, such that $X \succ Y$.

4 Results

The *entropy* of any discrete probability distribution $q = (q_1, \dots, q_K)$ is defined by $h(q) = \sum_{k=1}^K q_k \log_2 \left(\frac{1}{q_k} \right)$. (When $p = 0$, the term $p \log_2(1/p)$ is assigned the value $\lim_{p \rightarrow 0} [p \log_2(1/p)] = 0$.) The Mutual Information index equals the entropy of the district's ethnic distribution minus the average entropy of the ethnic distributions of its schools:

$$M(X) = h(P) - \sum_{n \in \mathbf{N}} P^n h(p^n)$$

where $P = (P_g)_{g \in \mathbf{G}}$ is the district ethnic distribution and $p^n = (p_g^n)_{g \in \mathbf{G}}$ is the ethnic distribution of school n . If the ethnic group and school of a randomly selected student are thought of as random variables \tilde{g} and \tilde{n} , then the Mutual Information equals the *mutual information* of these variables: the reduction in uncertainty about one variable that occurs when one learns the value of the other (Cover and Thomas [15, pp. 18 ff.]). Since mutual information is a symmetric concept,¹³

Observation 1 *the Mutual Information index is unchanged if the district matrix is transposed (i.e., relabeling ethnic groups as schools and vice-versa).*

Accordingly, the Mutual Information index captures the criteria of evenness and representativeness in a symmetric fashion. Also by symmetry, the Mutual Information index can be interpreted both as the information that a student's school conveys about her ethnicity, as well as what her ethnicity tells us about her school.

Our main result is that the segregation ordering represented by the Mutual Information index is the unique ordering that satisfies all of our axioms:

Theorem 1 *The Mutual Information ordering is the only segregation ordering that satisfies SYM, WSI, SDP, IND1, IND2, GDP, C, and N.*

¹³See Cover and Thomas [15] for a survey of this and other properties of mutual information.

The Mutual Information index is related to the more widely used Entropy segregation index (Theil [53]; Theil and Finizza [54]), which is given by

$$H(X) = 1 - \frac{\sum_{n \in \mathbf{N}} P^n h(p^n)}{h(P)} \quad (2)$$

The Entropy index is the result of dividing the Mutual Information index by its maximum value, the entropy h of the district ethnic distribution. Thus, the Entropy index takes a maximum value of one, while the Mutual Information index has no maximum value. These indices do not give the same segregation ordering. For instance, the Entropy index ranks all districts with no ethnic mixing as equally segregated, while the Mutual Information index assigns a higher segregation level to districts in which there is more initial uncertainty about a student's ethnicity.

Two examples illustrate this point. In the two districts $\langle(1, 0, 0), (0, 1, 0), (0, 0, 1)\rangle$ and $\langle(1, 0), (0, 1)\rangle$, learning a student's school uniquely determines her ethnicity. However, in the first district, there initially is more uncertainty about the student's ethnicity since there are three equal-size groups instead of two. According to the Mutual Information index, the first district is strictly more segregated ($M = 1.6$) than the second one ($M = 1.0$). The Entropy index treats these districts as equally segregated, assigning both an index of 1.0.

A student's ethnicity is determined by her school in the districts $X = \langle(99, 0), (0, 1)\rangle$ and $Y = \langle(50, 0), (0, 50)\rangle$ as well. But in the second district there initially is more uncertainty about a student's ethnicity than in the first. According to the Mutual Information index, the second district is more segregated ($M = 1.0$) than the first one ($M = 0.08$), these districts both have an Entropy index of 1.0.

In the context of school segregation, are normalized indices desirable? Clotfelter [7] argues not, on the grounds that they do not reflect changes in interracial contact well. To illustrate his point, consider merging the two schools in district X or district Y , defined in the prior paragraph. In district X , such a merger has a much smaller effect on the interracial exposure of the typical student, since 99% of students see only a 1% change in the percentage of minorities. The effects in Y are much greater, since 50% of each student's schoolmates

are now of the other racial group. While the Entropy index falls by the same amount, 1.0, in both cases, the Mutual Information index falls by 0.08 in district X versus 1.0 in Y .

5 Other Indices

This section presents and analyzes other indices that have been used in the literature on school segregation. In addition to our axioms, we also consider an additional property, Scale Invariance, and two decomposability properties.

Scale Invariance states that the segregation of a district is invariant to proportional changes in ethnic group size:¹⁴

Scale Invariance (SI) For any district X , ethnic group $g \in \mathbf{G}(X)$, and constant $\alpha > 0$, let X' be the result of multiplying the number of group- g students in each school n in district X by α . Then $X' \sim X$.

This property has both supporters and opponents in the field of school segregation (Taeuber and James [50, p. 134]; Coleman, Hoffer, and Kilgore [10, p. 178]).

The next property states that, for any partition of a district's schools into clusters, total segregation in the district is the sum of between-cluster and within-cluster segregation (which are defined in section 3):

Strong School Decomposability (SSD) An index S satisfies Strong School Decomposability if, for any partition $X = X^1 \uplus \dots \uplus X^K$ of the schools of a district into K clusters,

$$S(X) = S(c(X^1) \uplus \dots \uplus c(X^K)) + \sum_{k=1}^K P^k S(X^k) \quad (3)$$

where $S(c(X^1) \uplus \dots \uplus c(X^K))$ is segregation between the K clusters, $S(X^k)$ is segregation within cluster k , and P^k is the proportion of students in cluster k .

¹⁴This property is also known as Compositional Invariance (e.g., James and Taeuber [32, pp. 15-16]).

Mora and Ruiz-Castillo [39] show that the Mutual Information index satisfies SSD in the case of two groups. SSD and weaker forms of separability have also been extensively discussed in the literature of the measurement of income inequality. Bourguignon [4], for instance, shows that a property analogous to SSD fully characterizes the Theil inequality index (a close relative of the Mutual Information index) within the class of differentiable relative inequality indices. Foster [22] obtains a further characterization of the Theil inequality index by replacing the differentiability requirement by a more appealing transfer principle. Within the literature on the measurement of segregation, Hutchens [29] uses a weaker version of separability to help characterize a segregation index that represents the ordering induced by the Atkinson index (Atkinson [1]).

The second, analogous property states that, for any partition of a district’s groups into sets or “supergroups”, total segregation is the sum of between-supergroup and within-supergroup segregation (which are defined in section 3):

Strong Group Decomposability (SGD) An index S satisfies Strong Group Decomposability if, for any partition of the ethnic groups of a district X into K supergroups,

$$S = S_K + \sum_{k=1}^K P_k S_k \quad (4)$$

where S_K is segregation between the K supergroups, S_k is the segregation within supergroup k , and P_k is the proportion of students who are in supergroup k .

SSD and SGD have strong ordinal implications:

Proposition 1 *If S is a segregation index that satisfies Strong School Decomposability, then the segregation ordering represented by S satisfies IND1 and IND2. If S satisfies Strong Group Decomposability, then the induced segregation ordering satisfies GDP.*

A consequence is that if a segregation ordering does not satisfy GDP (respectively, either IND1 or IND2), then it cannot be represented by an index that satisfies SGD (respectively, SSD). The Mutual Information index is decomposable in both ways:

Proposition 2 *M satisfies SSD and SGD.*

It is easy to verify that the Mutual Information index does not satisfy SI. In the following claims, we state which of the ten properties SYM, WSI, SDP, IND1, IND2, GDP, C, N, SSD, and SGD are violated by other indices in the school segregation literature. Where we say that an index violates SSD or SGD, we also mean that the underlying ordering has no alternative representation that satisfies the given property. The proofs are straightforward and appear in an unpublished appendix (Frankel and Volij [21]). For proofs regarding SI, the reader is referred to Reardon and Firebaugh [45]. If a property is not mentioned, the index satisfies it.¹⁵ The notation I denotes the Simpson Interaction Index, $I = \sum_{g \in \mathbf{G}} P_g(1 - P_g)$ (Lieberman [35]).

5.1 Index of Dissimilarity

The Multigroup Dissimilarity Index D of Morgan [42] and Sakoda [47], a generalization of the 2-group index of Jahn, Schmid, and Schrag [30], is as follows:

$$D = \frac{1}{I}D' \text{ where } D' = \frac{1}{2} \sum_{g \in \mathbf{G}} \sum_{n \in \mathbf{N}} P^n |p_g^n - P_g| \text{ and}$$

Intuitively, D' equals the minimum proportion of the population that would have to change schools, keeping school sizes fixed, in order for each school to be representative of the district. I is what this proportion would be under complete segregation. Hence, the Multigroup Dissimilarity Index, D , is a normalization of D' that take a maximum value of 1.¹⁶ In the two-group case, the formula for D simplifies to $\frac{1}{2} \sum_{n \in \mathbf{N}} \left| \frac{T_1^n}{T_1} - \frac{T_2^n}{T_2} \right|$.

Claim 1 *The multigroup dissimilarity index D satisfies all properties but IND1, IND2, GDP, SSD, and SGD. It satisfies SI only in the two-group case.*

¹⁵More precisely, the index satisfies the restriction of the property to cities with at least two nonempty groups. Unlike the Mutual Information index, the indices in this section are not defined on the (albeit uninteresting) set of cities that contain only one nonempty group.

¹⁶Some researchers (e.g., Watts [55], in the case of occupational gender segregation) have instead used the unnormalized version, D' .

5.2 Gini

The multigroup Gini index of Reardon [44], is a generalization of the two-group Gini index of Jahn, Schmidt, and Schrag [30]:

$$\mathcal{G} = \frac{1}{2I} \sum_{g \in \mathbf{G}} \sum_{m \in \mathbf{N}} \sum_{n \in \mathbf{N}} P^m P^n |p_g^m - p_g^n|$$

Claim 2 *The multigroup Gini index \mathcal{G} satisfies all properties but IND1, IND2, GDP, SSD, and SGD. It satisfies SI only in the two-group case.*

5.3 Entropy Index

The Entropy index is defined in equation (2).

Claim 3 *The Entropy index H satisfies all properties but IND2, GDP, SI, SSD, and SGD.*

5.4 Normalized Exposure

The Normalized Exposure index was originally proposed by Bell [2] for the case of two groups. Its multigroup version, formulated by James [31], is

$$P = \sum_{g \in \mathbf{G}} \sum_{n \in \mathbf{N}} P^n \frac{(p_g^n - P_g)^2}{1 - P_g}$$

In the case of two groups (say whites and blacks, denoted 1 and 2, respectively), the index equals $\frac{P_2 - E^*}{P_2}$ where $E^* = \frac{1}{T_1} \sum_{n \in \mathbf{N}} T_1^n p_2^n$ is the “exposure” of whites to blacks: the proportion black in the school attended by the average white student, and P_2 (the proportion black in the district) is the maximum value E^* can take.¹⁷ Thus, the two-group index measures the exposure of whites to blacks, normalized by the maximum possible such exposure. The index is symmetric: it also measures the normalized exposure of blacks to whites.

¹⁷Proof available on request.

Claim 4 *The Normalized Exposure index P satisfies all properties but GDP, SI, IND2, SSD, and SGD. It satisfies IND1 only in the two-group case.*

5.5 Other Measures of School Segregation

Research that relies on the above indices includes Reardon and Yun [46] and Taeuber and James [50], both of whom use D , G , and H , and Zoloth [56], who uses D , P , and H . In addition, P is used by Coleman, Kelly, and Moore [11] and Coleman, Hoffer, and Kilgore [10]. Other research on school segregation relies on different measures. One is the percentage of blacks or nonwhites who attend schools in which at least some proportion κ of students are nonwhite. If we simplify by assuming two groups, whites (group 1) and blacks (group 2), this index can be written

$$Cl(X) = \frac{1}{T_2} \sum_{n \in \mathbf{N}(X): p_2^n \geq \kappa} T_2^n \quad (5)$$

This measure is also used by Clotfelter [8] (who also uses P , D , and G), Clotfelter, Ladd, and Vigdor [9] (who also use P) and Boozer, Krueger, and Wolkon [3].

Claim 5 *The index Cl violates SYM, SDP, IND2, SI, and SSD. It satisfies WSI, N, IND1, and C.*

There is ambiguity regarding how to generalize Cl to an arbitrary number of groups, so we cannot say whether it satisfies GDP or SGD.

Card and Rothstein [5] compute the average fraction black or Hispanic in the schools attended by the typical black and white student, and define segregation as the difference between these figures. Letting whites, blacks, and Hispanics be indexed by 1, 2, and 3, respectively, this index equals

$$CR(X) = \sum_{n \in \mathbf{N}(X)} \left(\frac{T_2^n}{T_2} - \frac{T_1^n}{T_1} \right) \frac{T_2^n + T_3^n}{T^n}$$

Claim 6 *The index CR violates SYM, SDP, IND1, IND2, SI, and SSD. It satisfies WSI, N, and C.*

Since CR is defined for three particular groups, GDP and SGD cannot be evaluated.

5.6 Summary of Results

The results of this section are summarized in the following table. A check mark indicates that the property is satisfied; “ \times ” indicates that it is violated.

| Index | SYM | WSI | SDP | IND1 | IND2 | GDP | C | N | SI | SSD | SGD |
|---------------------|----------|-----|----------|----------|----------|----------|---|---|----------|----------|----------|
| Modified Entropy | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | \times | ✓ | ✓ |
| Entropy | ✓ | ✓ | ✓ | ✓ | \times | \times | ✓ | ✓ | \times | \times | \times |
| Dissimilarity | ✓ | ✓ | ✓ | \times | \times | \times | ✓ | ✓ | 2 | \times | \times |
| Gini | ✓ | ✓ | ✓ | \times | \times | \times | ✓ | ✓ | 2 | \times | \times |
| Normalized Exposure | ✓ | ✓ | ✓ | 2 | \times | \times | ✓ | ✓ | \times | \times | \times |
| CI | \times | ✓ | \times | ✓ | \times | N/A | ✓ | ✓ | \times | \times | N/A |
| CR | \times | ✓ | \times | \times | \times | N/A | ✓ | ✓ | \times | \times | N/A |

Table 2: Properties of School Segregation Indices. A check mark indicates that the property is satisfied by the index. An “ \times ” indicates that it is not. “2” indicates that it is satisfied only in the 2-group case. The properties are Symmetry (SYM), Weak Scale Invariance (WSI), the School Division Property (SDP), Type I Independence (IND1), Type II Independence (IND2), the Group Division Property (GDP), Continuity (C), Nontriviality (N), Scale Invariance (SI), Strong School Decomposability (SSD), and Strong Group Decomposability (SGD).

6 Related Literature

The first to study segregation axiomatically was Philipson [43], who provides an axiomatic characterization of a large family of segregation orderings that have an additively separable representation. The representation consists of a weighted average of a function that depends on the school’s ethnic distribution only.

In two papers, Hutchens [28, 29] studies the measurement of segregation in the case of two ethnic groups. Hutchens [28] characterizes the family of indices that satisfy a set of mostly cardinal properties. Hutchens [29] strengthens one axiom and obtains a unique segregation index, which is based on the Atkinson inequality index [1]. Both of these papers assume Scale Invariance.

Echenique and Fryer [18] use data on individuals' social networks to measure the strength of an individual's isolation from members of other ethnic groups. They rely on cardinal axioms and require data on social networks, while our measure relies only on ordinal axioms and uses data on the numbers of students in each ethnic group and school.

A Proofs

Proof of Theorem 1. We first show that the ordering represented by the Mutual Information index satisfies the axioms. Axioms N, SYM, and WSI are trivial, and C follows from the fact that the index M is a continuous function of the T_g^n 's (the number of students of each group in each school). Axioms IND1, IND2, and GDP follow from Propositions 1 and 2. So it remains to show that SDP is satisfied. Let $X \in \mathcal{C}$ be a district and let n be a school of X . Let X' be the district that results from dividing n into two schools, n_1 and n_2 . Since X and X' have the same group distribution,

$$\begin{aligned} M(X') - M(X) &= P^n h((p_g^n)_{g \in \mathbf{G}(X)}) - P^{n_1} h((p_g^{n_1})_{g \in \mathbf{G}(X)}) - P^{n_2} h((p_g^{n_2})_{g \in \mathbf{G}(X)}) \\ &= P^n \left(h((p_g^n)_{g \in \mathbf{G}(X)}) - \frac{P^{n_1}}{P^n} h((p_g^{n_1})_{g \in \mathbf{G}(X)}) - \frac{P^{n_2}}{P^n} h((p_g^{n_2})_{g \in \mathbf{G}(X)}) \right) \end{aligned}$$

But for all g , $p_g^n = \frac{P^{n_1}}{P^n} p_g^{n_1} + \frac{P^{n_2}}{P^n} p_g^{n_2}$ so, recalling that $h((q_g)_{g \in \mathbf{G}}) = \sum_{g \in \mathbf{G}} q_g \log_2(\frac{1}{q_g})$ is a concave function, $M(X') - M(X) \geq 0$, with strict inequality only if schools n_1 and n_2 have different group distributions. This verifies SDP.

We now show that the Mutual Information ordering is the only segregation ordering that satisfies all the axioms. Let \succsim be a segregation ordering that satisfies them. For any district X , let the schools be numbered $n = 1, \dots, N$ and the groups $g = 1, \dots, G$.

For any group distribution $P = (P_g)_{g=1}^G$, let $\overline{X}(P)$ denote the district, with population 1, with group distribution P , and with G uniracial schools, and let $\underline{X}(P)$ denote the one-school district with group distribution P and population 1:

$$\overline{X}(P) = \langle (P_1, 0, \dots, 0), \dots, (0, \dots, 0, P_G) \rangle \quad \text{and} \quad \underline{X}(P) = \langle (P_1, \dots, P_G) \rangle.$$

For any integer $G \geq 1$, let $\overline{X}^G = \langle (1/G, 0, \dots, 0), \dots, (0, \dots, 0, 1/G) \rangle$ denote the completely segregated district of population 1 with G equal sized ethnic groups. Let $\underline{X}^G = \langle (1/G, \dots, 1/G) \rangle$ denote the one-school district with the same group distribution and population.

We first state and prove some preliminary lemmas. By applying IND1 repeatedly, one can show the following apparently stronger (but actually equivalent) property, which will be used interchangeably with IND1.

Lemma 1 *Suppose the segregation ordering \succsim satisfies IND1. Let $X, Y \in C$ be two districts with equal populations and equal group distributions. Then for all districts $Z \in C$ containing any number of schools, $X \succsim Y$ if and only if $X \uplus Z \succsim Y \uplus Z$.*

Proof. Let the schools of Z be enumerated: n_1, \dots, n_N . By IND1, $X \succsim Y$ if and only if $X \uplus \langle n_1 \rangle \succsim Y \uplus \langle n_1 \rangle$, where $\langle n_1 \rangle$ denotes a district that consists of school n_1 alone. The districts $X' = X \uplus \langle n_1 \rangle$ and $Y' = Y \uplus \langle n_1 \rangle$ have the same size and group distribution since X and Y do. Hence, by IND1, $X' \succsim Y'$ if and only if $X' \uplus \langle n_2 \rangle \succsim Y' \uplus \langle n_2 \rangle$. The result follows by repeating the same argument for schools n_3, \dots, n_N . Q.E.D.

Lemma 2 *1. All districts in which every school is representative have the same degree of segregation under \succsim .*

2. Any district in which every school is representative is weakly less segregated under \succsim than any district in which some school is unrepresentative.

Proof.

1. Consider any district Y that consists of N representative schools. By WSI we can assume w.l.o.g. that $T(Y) = 1$. For each $i = 1, \dots, N$, let Y_i be the school district consisting of schools $i + 1$ through N of Y as well as a single school that contains the students in schools 1 through i of Y . By SDP, for each $i = 1, \dots, N - 1$, $Y_i \sim Y_{i+1}$. Hence, by transitivity, $Y = Y_1 \sim Y_N$. Y_N contains a single school. By GDP, $Y_N \sim \underline{X}^1$, and hence $Y \sim \underline{X}^1$.

2. Consider any district Y in which at least one school is unrepresentative. The above procedure yields $Y = Y_1 \succcurlyeq Y_2 \succcurlyeq \dots \succcurlyeq Y_N \sim \underline{X}^1$. By transitivity, $Y \succcurlyeq \underline{X}^1$.

Q.E.D.

Lemma 3 *For any district Z with G ethnic groups, let $\sigma(Z) \in \mathcal{C}$ be such that the number of persons of ethnic group g in school n in Z equals the number of persons of ethnic group $(g + 1) \bmod G$ in school n in $\sigma(Z)$. Define $\sigma^1(Z) = \sigma(Z)$ and, for integers $j > 1$, let $\sigma^j(Z) = \sigma(\sigma^{j-1}(Z))$.¹⁸ Then $\frac{1}{G} \uplus_{j=1}^G \sigma^j(Z) \succcurlyeq Z$.*

Proof. Consider the following statement:

$$\left(\uplus_{j=1}^n Z \right) \uplus \left(\uplus_{j=n+1}^G c(Z) \right) \preccurlyeq \left(\uplus_{j=1}^n \sigma^j(Z) \right) \uplus \left(\uplus_{j=n+1}^G \sigma^j(c(Z)) \right) \quad (6)$$

For $n = 0$, (6) simply states $\uplus_{j=1}^G c(Z) \preccurlyeq \left(\uplus_{j=1}^G \sigma^j(c(Z)) \right)$, which holds by Lemma 2. Assume that (6) holds for some $n = k$, with $0 \leq k < G - 1$. Then, taking into account that σ^G is the identity permutation,

$$\begin{aligned} & \left(\uplus_{j=1}^n Z \right) \uplus \left(\uplus_{j=n+2}^G c(Z) \right) \uplus c(Z) \preccurlyeq \left(\uplus_{j=1}^n \sigma^j(Z) \right) \uplus \left(\uplus_{j=n+1}^{G-1} \sigma^j(c(Z)) \right) \uplus c(Z) \\ \implies & \left(\uplus_{j=1}^n Z \right) \uplus \left(\uplus_{j=n+2}^G c(Z) \right) \uplus Z \preccurlyeq \left(\uplus_{j=1}^n \sigma^j(Z) \right) \uplus \left(\uplus_{j=n+1}^{G-1} \sigma^j(c(Z)) \right) \uplus Z \quad \text{by IND2} \\ & \sim \sigma \left(\left(\uplus_{j=1}^n \sigma^j(Z) \right) \uplus \left(\uplus_{j=n+1}^{G-1} \sigma^j(c(Z)) \right) \uplus Z \right) \quad \text{by SYM} \\ & \sim \left(\uplus_{j=2}^{n+1} \sigma^j(Z) \right) \uplus \left(\uplus_{j=n+2}^G \sigma^j(c(Z)) \right) \uplus \sigma(Z) \quad \text{by def. of } \sigma \\ & \left(\uplus_{j=1}^{n+1} Z \right) \uplus \left(\uplus_{j=n+2}^G c(Z) \right) \preccurlyeq \left(\uplus_{j=1}^{n+1} \sigma^j(Z) \right) \uplus \left(\uplus_{j=n+2}^G \sigma^j(c(Z)) \right) \end{aligned}$$

That is, (6) also holds for $n = k + 1$. By induction it also holds for $n = G - 1$. That is, $\uplus_{j=1}^G Z \preccurlyeq \uplus_{j=1}^G \sigma^j(Z)$ which, by SDP and WSI implies $Z \preccurlyeq \frac{1}{G} \uplus_{j=1}^G \sigma^j(Z)$. Q.E.D.

Lemma 4 *For any district X with G groups and group distribution P , $\overline{X}^G \succcurlyeq \overline{X}(P) \succcurlyeq X$.*

¹⁸Note that $\sigma^G(Z) = Z$.

Proof. By WSI, w.l.o.g. we can assume that $T(X) = 1$. X can be converted into a completely segregated district by dividing each school n into G distinct schools, each of which includes all and only the members of a single ethnic group. By SDP, this procedure results in a weakly more segregated district. By then combining all schools containing a given ethnic group, this can be converted to $\overline{X}(P)$ without changing the segregation level (by SDP). To see that $\overline{X}^G \succcurlyeq X$, note that by Lemma 3, $\frac{1}{G} \uplus_{j=1}^G \sigma^j(\overline{X}(P)) \succcurlyeq \overline{X}(P)$. But by SDP, the left hand side district is as segregated as \overline{X}^G . Q.E.D.

Lemma 5 *For any integer $G \geq 1$, $\overline{X}^G \preccurlyeq \overline{X}^{G+1}$.*

Proof. Let X be the $(G + 1)$ -group district that results after splitting one ethnic group in \overline{X}^G up into two equally distributed subgroups. By Lemma 4 and GDP, $\overline{X}^{G+1} \succcurlyeq X \sim \overline{X}^G$. Q.E.D.

Lemma 6 *Let X and X' be two districts with the same size and group distribution such that $X \succ X'$. Let $1 \geq \alpha > \beta \geq 0$. Then $\alpha X \uplus (1 - \alpha)X' \succ \beta X \uplus (1 - \beta)X'$*

Proof. By WSI, $(\alpha - \beta)X \succ (\alpha - \beta)X'$. Since X and X' have the same size and group distribution, so do $(\alpha - \beta)X$ and $(\alpha - \beta)X'$. So by IND1,

$$\beta X \uplus (\alpha - \beta)X \uplus (1 - \alpha)X' \succ \beta X \uplus (\alpha - \beta)X' \uplus (1 - \alpha)X'.$$

By SDP, $\alpha X \uplus (1 - \alpha)X' \succ \beta X \uplus (1 - \beta)X'$. Q.E.D.

Lemma 7 *For any districts $Z \succcurlyeq X \succcurlyeq Y$ such that $Z \succ Y$ and Y and Z have the same size and group distribution, there is a unique $\alpha \in [0, 1]$ such that $X \sim \alpha Z \uplus (1 - \alpha)Y$.*

Proof. The sets $\{\alpha \in [0, 1] : \alpha Z \uplus (1 - \alpha)Y \succcurlyeq X\}$ and $\{\alpha \in [0, 1] : X \succcurlyeq \alpha Z \uplus (1 - \alpha)Y\}$ are closed by C. Any α satisfies $X \sim \alpha Z \uplus (1 - \alpha)Y$ if and only if it is in the intersection of these two sets. Given that $Z \succcurlyeq X \succcurlyeq Y$, these sets are each nonempty. Their union is the whole unit interval since \succcurlyeq is complete. Since the interval $[0, 1]$ is connected, the intersection

of the two sets must be nonempty. By Lemma 6, their intersection cannot contain more than one element. Thus, their intersection contains a single element α . Q.E.D.

Let X be a district with G groups and group distribution $\hat{P} = (\hat{P}_1, \dots, \hat{P}_G)$. For any $G' \geq 1$ and any distribution $P = (P_1, \dots, P_{G'})$ let $\phi^P(X)$ be the district that results after splitting each ethnic group g in district X into G' ethnic groups in proportions given by P . That is, the T_g^n members of each ethnic group g in each school n of X are split up into G' ethnic groups of size $P_1 T_g^n, \dots, P_{G'} T_g^n$. The resulting district $\phi^P(X)$ has GG' groups with distribution $\left((\hat{P}_g P_{g'})_{g=1}^G \right)_{g'=1}^{G'}$.

Let X be a district and let $\hat{P} = (\hat{P}_1, \dots, \hat{P}_G)$ be an arbitrary distribution such that $\overline{X}(\hat{P}) \succcurlyeq X$ and $\overline{X}(\hat{P}) \succcurlyeq \overline{X}^2$. By lemmas 4 and 5 such a distribution exists. By Nontriviality, Lemma 4, and Lemma 2, $\overline{X}^2 \succ \underline{X}^2 \sim \underline{X}(\hat{P})$. Therefore, by Lemma 7 there is a unique $\hat{\alpha}$ such that

$$X \sim \hat{\alpha} \overline{X}(\hat{P}) \uplus (1 - \hat{\alpha}) \underline{X}(\hat{P}). \quad (7)$$

Similarly, by Lemma 7 there is a unique $\hat{\beta}$ such that $\overline{X}^2 \sim \hat{\beta} \overline{X}(\hat{P}) \uplus (1 - \hat{\beta}) \underline{X}(\hat{P})$. By Lemma 6, $\hat{\beta} > 0$, as $\overline{X}(\hat{P}) \succcurlyeq \overline{X}^2$.

Define the index $S : \mathcal{C} \rightarrow \mathbb{R}$ by

$$S(X) = \hat{\alpha} / \hat{\beta} \quad (8)$$

For S to be well defined, $\hat{\alpha} / \hat{\beta}$ cannot depend on the particular choice of \hat{P} . We now verify this. Consider another distribution $\tilde{P} = (\tilde{P}_1, \dots, \tilde{P}_{G'})$ such that $\overline{X}(\tilde{P}) \succcurlyeq X$ and $\overline{X}(\tilde{P}) \succcurlyeq \overline{X}^2$ and let $\tilde{\alpha}$ and $\tilde{\beta}$ the unique numbers such that $X \sim \tilde{\alpha} \overline{X}(\tilde{P}) \uplus (1 - \tilde{\alpha}) \underline{X}(\tilde{P})$ and $\overline{X}^2 \sim \tilde{\beta} \overline{X}(\tilde{P}) \uplus (1 - \tilde{\beta}) \underline{X}(\tilde{P})$. By GDP

$$X \sim \phi^{\hat{P}} \left(\tilde{\alpha} \overline{X}(\tilde{P}) \uplus (1 - \tilde{\alpha}) \underline{X}(\tilde{P}) \right) \sim \tilde{\alpha} \phi^{\hat{P}} \left(\overline{X}(\tilde{P}) \right) \uplus (1 - \tilde{\alpha}) \phi^{\hat{P}} \left(\underline{X}(\tilde{P}) \right) \quad (9)$$

Similarly, applying the transformation $\phi^{\tilde{P}}$ to (7) and using GDP,

$$X \sim \hat{\alpha} \phi^{\tilde{P}} \left(\overline{X}(\hat{P}) \right) \uplus (1 - \hat{\alpha}) \phi^{\tilde{P}} \left(\underline{X}(\hat{P}) \right) \quad (10)$$

Both $\phi^{\hat{P}}(\underline{X}(\hat{P}))$ and $\phi^{\hat{P}}(\underline{X}(\tilde{P}))$ are districts with the same number of groups ($G * G'$) and (up to a permutation) the same group distribution. Further by SYM, $\phi^{\tilde{P}}(\underline{X}(\hat{P})) \sim \phi^{\hat{P}}(\underline{X}(\tilde{P}))$. Similarly, both $\phi^{\tilde{P}}(\overline{X}(\hat{P}))$ and $\phi^{\tilde{P}}(\overline{X}(\tilde{P}))$ are districts with the same number of groups and (up to a permutation) the same group distribution. Assume w.l.o.g. that $\phi^{\hat{P}}(\overline{X}(\tilde{P})) \succcurlyeq \phi^{\tilde{P}}(\overline{X}(\hat{P}))$ and let γ be the unique number such that

$$\phi^{\tilde{P}}(\overline{X}(\hat{P})) \sim \gamma \phi^{\hat{P}}(\overline{X}(\tilde{P})) \uplus (1 - \gamma) \phi^{\hat{P}}(\underline{X}(\tilde{P}))$$

Then, applying WSI, IND1 (twice) and SDP, it follows from (10) that

$$\begin{aligned} X &\sim \hat{\alpha} \left[\gamma \phi^{\hat{P}}(\overline{X}(\tilde{P})) \uplus (1 - \gamma) \phi^{\hat{P}}(\underline{X}(\tilde{P})) \right] \uplus (1 - \hat{\alpha}) \phi^{\hat{P}}(\underline{X}(\tilde{P})) \\ &\sim \hat{\alpha} \gamma \phi^{\hat{P}}(\overline{X}(\tilde{P})) \uplus (1 - \gamma \hat{\alpha}) \phi^{\hat{P}}(\underline{X}(\tilde{P})) \end{aligned} \quad (11)$$

Comparing (11) and (9) we obtain that $\tilde{\alpha} = \hat{\alpha} \gamma$. Exactly the same reasoning leads to $\tilde{\beta} = \hat{\beta} \gamma$. Consequently $\hat{\alpha} / \hat{\beta} = \tilde{\alpha} / \tilde{\beta}$. This establishes that S is well-defined.

Lemma 8 *The index S defined in (8) represents \succcurlyeq .*

Proof. Let $X, Y \in \mathcal{C}$ and let G be at least as large as the number of groups in X or Y . Then, by lemmas 4 and 5, $\overline{X}^G \succcurlyeq \overline{X}^2$, $\overline{X}^G \succcurlyeq X$ and $\overline{X}^G \succcurlyeq Y$. Define α_X , α_Y and β by

$$\begin{aligned} X &\sim \alpha_X \overline{X}^G \uplus (1 - \alpha_X) \underline{X}^G \\ Y &\sim \alpha_Y \overline{X}^G \uplus (1 - \alpha_Y) \underline{X}^G \\ \overline{X}^2 &\sim \beta \overline{X}^G \uplus (1 - \beta) \underline{X}^G. \end{aligned}$$

Then,

$$\begin{aligned}
X \succcurlyeq Y &\iff \alpha_X \overline{X}^G \uplus (1 - \alpha_X) \underline{X}^G \succcurlyeq \alpha_Y \overline{X}^G \uplus (1 - \alpha_Y) \underline{X}^G && \text{by definition of } \alpha_X \text{ and } \alpha_Y \\
&\iff \alpha_X \geq \alpha_Y && \text{by Lemma 6} \\
&\iff \alpha_X / \beta \geq \alpha_Y / \beta && \text{since } \beta > 0 \\
&\iff S(X) \geq S(Y) && \text{by definition of } S
\end{aligned}$$

Q.E.D.

The following results will be used to show that S is the Mutual Information index.

Lemma 9 *For any group distribution $P = (P_1, \dots, P_G)$ (in which some entries may be zero), let $\hat{P} = (\frac{P_1}{G}, \dots, \frac{P_1}{G}, \dots, \frac{P_G}{G}, \dots, \frac{P_G}{G})$ be the group distribution that results from dividing each ethnic group in P into G equal sized groups. Then $\overline{X}(\hat{P}) \succcurlyeq \overline{X}^G$ and $\overline{X}(\hat{P}) \succcurlyeq \overline{X}(P)$.*

Proof. For the first claim, first subdivide each ethnic group in \overline{X}^G into G groups in proportions given by P . For instance, the first group is divided into G groups of sizes $P_1 \frac{1}{G}, \dots, P_G \frac{1}{G}$. Now put each resulting group in a separate school. The group distribution of the resulting district, $(P_1 \frac{1}{G}, \dots, P_G \frac{1}{G}, \dots, P_1 \frac{1}{G}, \dots, P_G \frac{1}{G})$, is just a permutation of \hat{P} . Hence, by GDP and SDP, $\overline{X}(\hat{P}) \succcurlyeq \overline{X}^G$. The second claim follows from the first one after noting that by Lemma 4, $\overline{X}^G \succcurlyeq \overline{X}(P)$. Q.E.D.

Lemma 10 *Let districts $Z_1, Z_2, Z_3,$ and Z_4 all have the same population and group distribution and let $Z_1 \sim Z_2$ and $Z_3 \sim Z_4$. Let Z_5, Z_6 be two districts with equal populations. Then $Z_1 \uplus Z_5 \sim Z_2 \uplus Z_6$ if and only if $Z_3 \uplus Z_5 \sim Z_4 \uplus Z_6$.*

Proof. By IND2 applied twice, $Z_1 \uplus Z_5 \sim Z_1 \uplus Z_6$ if and only if $Z_3 \uplus Z_5 \sim Z_3 \uplus Z_6$. But by IND1, $Z_1 \uplus Z_6 \sim Z_2 \uplus Z_6$ and $Z_3 \uplus Z_6 \sim Z_4 \uplus Z_6$. Q.E.D.

Lemma 11 *For any districts X and Y , $S(X \uplus Y) = S(c(X) \uplus Y) + \frac{T(X)}{T(X)+T(Y)} S(X)$.*

Proof. Let X and Y be any two districts. Let $X \uplus Y$ have G ethnic groups. By adding an empty group if needed, we can assume WLOG that $G \geq 2$. For any district

Z , let $\phi^G(Z)$ be the result of splitting each group g in Z into G equal-size groups, each of which has the same school distribution as g . Let \hat{P} be the group distribution of $\phi^G(X)$. By Lemma 9, $\overline{X}(\hat{P}) \succcurlyeq \overline{X}^G$. Define $\hat{\alpha}_X$ by $X \sim \hat{\alpha}_X \overline{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X) \underline{X}(\hat{P})$ and γ by $c(X) \uplus Y \sim \gamma \overline{X}(\hat{P}) \uplus (1 - \gamma) \underline{X}(\hat{P})$. Define

$$\begin{aligned}
Z_1 &= \phi^G(X) \\
Z_2 &= T(X) \left(\hat{\alpha}_X \overline{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X) \underline{X}(\hat{P}) \right) \\
Z_3 &= c(\phi^G(X)) = \phi^G(c(X)) \\
Z_4 &= T(X) \underline{X}(\hat{P}) \\
Z_5 &= \phi^G(Y) \\
Z_6 &= T(X \uplus Y) \left(\gamma \overline{X}(\hat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma \right) \underline{X}(\hat{P}) \right)
\end{aligned}$$

To show that Z_6 is well defined, we must show that $\gamma \leq 1 - \frac{T(X)}{T(X \uplus Y)} = \frac{T(Y)}{T(X \uplus Y)}$. For this, by Lemma 6, it is enough to show that

$$\gamma \overline{X}(\hat{P}) \uplus (1 - \gamma) \underline{X}(\hat{P}) \preccurlyeq \frac{T(Y)}{T(X \uplus Y)} \overline{X}(\hat{P}) \uplus \frac{T(X)}{T(X \uplus Y)} \underline{X}(\hat{P}). \quad (12)$$

The district $c(X) \uplus Y$ has G groups since $X \uplus Y$ does. By Lemma 3,

$$c(X) \uplus Y \preccurlyeq \frac{1}{G} \uplus_{j=1}^G \sigma^j(c(X) \uplus Y) = \frac{1}{G} \uplus_{j=1}^G \sigma^j(c(X)) \uplus \frac{1}{G} \uplus_{j=1}^G \sigma^j(Y)$$

Let $\widetilde{c(X)} = \frac{1}{G} \uplus_{j=1}^G \sigma^j(c(X))$ and $\tilde{Y} = \frac{1}{G} \uplus_{j=1}^G \sigma^j(Y)$. Each of $\widetilde{c(X)}$ and \tilde{Y} has G groups of equal size. By SDP, $\widetilde{c(X)} \sim T(X) \underline{X}^G$ and both of these districts have the same population, $T(X)$, and the same group distribution. Since \tilde{Y} has G equal size groups, it is not more segregated than $T(Y) \overline{X}^G$ and both of these districts have the same population, $T(Y)$, and

the same group distribution. Therefore,

$$\begin{aligned}
c(X) \uplus Y &\preceq T(X)\underline{X}^G \uplus T(Y)\overline{X}^G && \text{by IND1 (twice)} \\
&\sim \phi^G \left(T(X)\underline{X}^G \uplus T(Y)\overline{X}^G \right) && \text{by GDP} \\
&\sim T(X)\phi^G(\underline{X}^G) \uplus T(Y)\phi^G(\overline{X}^G) && \text{by definition of } \phi^G \\
&\preceq T(X)\underline{X}(\hat{P}) \uplus T(Y)\overline{X}(\hat{P}) && \text{by SDP} \\
&\sim \frac{T(X)}{T(X \uplus Y)}\underline{X}(\hat{P}) \uplus \frac{T(Y)}{T(X \uplus Y)}\overline{X}(\hat{P}) && \text{by WSI}
\end{aligned}$$

But $c(X) \uplus Y \sim \gamma\overline{X}(\hat{P}) \uplus (1 - \gamma)\underline{X}(\hat{P})$ so (12) holds and $\gamma \leq \frac{T(Y)}{T(X \uplus Y)}$, as claimed.

By construction, $Z_1, Z_2, Z_3,$ and Z_4 all have the same population and group distribution. By GDP, $Z_1 \sim Z_2$. Clearly, $Z_3 \sim Z_4$ since these are actually the same district. Also, the population of Z_6 is $T(Y)$, which equals the population of Z_5 . Moreover,

$$\begin{aligned}
Z_4 \uplus Z_6 &= T(X)\underline{X}(\hat{P}) \uplus T(X \uplus Y) \left(\gamma\overline{X}(\hat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma \right) \underline{X}(\hat{P}) \right) \\
&= T(X \uplus Y) \left(\gamma\overline{X}(\hat{P}) \uplus (1 - \gamma)\underline{X}(\hat{P}) \right) && \text{by SDP} \\
&\sim c(X) \uplus Y && \text{by WSI} \\
&\sim \phi^G(c(X)) \uplus \phi^G(Y) && \text{by GDP} \\
&= Z_3 \uplus Z_5
\end{aligned}$$

So by Lemma 10,

$$Z_1 \uplus Z_5 \sim Z_2 \uplus Z_6 \tag{13}$$

Now,

$$\begin{aligned}
X \uplus Y &\sim \phi^G(X \uplus Y) \text{ by GDP} \\
&= Z_1 \uplus Z_5 \\
&\sim Z_2 \uplus Z_6 \text{ by (13)} \\
&= T(X) \left(\hat{\alpha}_X \bar{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X) \underline{X}(\hat{P}) \right) \\
&\quad \uplus T(X \uplus Y) \left(\gamma \bar{X}(\hat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma \right) \underline{X}(\hat{P}) \right) \\
&\sim (T(X \uplus Y)\gamma + T(X)\hat{\alpha}_X) \bar{X}(\hat{P}) \uplus T(X \uplus Y) \left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)} \hat{\alpha}_X \right) \underline{X}(\hat{P}) \text{ by SDP} \\
&\sim \left(\gamma + \frac{T(X)}{T(X \uplus Y)} \hat{\alpha}_X \right) \bar{X}(\hat{P}) \uplus \left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)} \hat{\alpha}_X \right) \underline{X}(\hat{P}) \text{ by WSI.}
\end{aligned}$$

We have shown that $X \uplus Y \sim \left(\gamma + \frac{T(X)}{T(X \uplus Y)} \hat{\alpha}_X \right) \bar{X}(\hat{P}) \uplus \left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)} \hat{\alpha}_X \right) \underline{X}(\hat{P})$. By definition of γ and $\hat{\alpha}_X$, $c(X) \uplus Y \sim \gamma \bar{X}(\hat{P}) \uplus (1 - \gamma) \underline{X}(\hat{P})$ and $X \sim \hat{\alpha}_X \bar{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X) \underline{X}(\hat{P})$. By Lemma 7, there is a unique β such that $\bar{X}^2 \sim \beta \bar{X}(\hat{P}) \uplus (1 - \beta) \underline{X}(\hat{P})$. By definition of S , $S(X \uplus Y) = \frac{1}{\beta} \left(\gamma + \frac{T}{T+T(Y)} \hat{\alpha}_X \right) = S(c(X) \uplus Y) + \frac{T}{T+T(Y)} S(X)$, as claimed. Q.E.D.

For any discrete probability distribution $P = (P_1, \dots, P_G)$, define the function $s(P)$ to equal $S(\bar{X}(P))$.

Claim 7 *The function s is the entropy function. Namely, $s(P) = h(P) = \sum_{i=1}^n P_i \log_2 \frac{1}{P_i}$.*

Proof. It is known that the entropy function is the only function that satisfies the following three properties.¹⁹

1. $h(1/2, 1/2) = 1$.
2. $h(p, 1 - p)$ is continuous in p .
3. $h(p_1, \dots, p_n) = h(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) h\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$.

¹⁹The statement of this result appears as an exercise in Cover and Thomas [15]. For the original proof, see Faddeev [19].

So it is enough to show that s satisfies them. Property 1 follows from the definition of S and the fact that $S(\overline{X}(1/2, 1/2)) = S(\overline{X}^2)$. Property 3 follows from Lemma 11. It remains to show property 2. Let us write $\overline{X}(p, 1-p)$ as Z^p for brevity. By Lemma 7, there is a unique α_p such that $Z^p \sim \alpha_p \overline{X}^2 \uplus (1 - \alpha_p) \underline{X}^2$. By definition of S , $\alpha_p = S(\overline{X}(p, 1-p))$. For all p , the sets $\{q : Z^q \succcurlyeq Z^p\}$ and $\{q : Z^q \preccurlyeq Z^p\}$ are closed by Continuity. Note that $Z^q \succcurlyeq Z^p$ if and only if $\alpha_q \geq \alpha_p$ by Lemma 6. So the sets $\{q : \alpha_q \geq \alpha_p\}$ and $\{q : \alpha_q \leq \alpha_p\}$ are closed. If α_p is not a continuous function of p , then let the sequence $(p_k)_{k=1}^\infty$ converge to some p . By restricting to an appropriate subsequence, we may assume that $\lim_{k \rightarrow \infty} \alpha_{p_k}$ exists. Let this limit be c and assume by contradiction that $c \neq \alpha_p$. Assume that $c > \alpha_p$ (the other case is analogous). Since $\lim_{k \rightarrow \infty} \alpha_{p_k} = c > \frac{c + \alpha_p}{2}$, there is a k^* such that $\alpha_{p_k} > \frac{c + \alpha_p}{2}$ for all $k > k^*$. So the sequence $\{p_k : k > k^*\}$ lies in $\{q : \alpha_q \geq \frac{c + \alpha_p}{2}\}$. But $\lim_{k \rightarrow \infty} p_k = p$ does not lie in this set, which contradicts the fact that this set is closed. Q.E.D.

We now show that S is the Mutual Information index. Consider any district X with N schools, G ethnic groups, and group distribution P . Let $X_0 = X$. Let X_n be the result of separating the students in each school $m \leq n$ into G uniracial schools. For instance, if $X = \langle (1, 2), (3, 4) \rangle$, then $X_1 = \langle (1, 0), (0, 2), (3, 4) \rangle$ and $X_2 = \langle (1, 0), (0, 2), (3, 0), (0, 4) \rangle$. Note that X_N is completely segregated and has group distribution P , so $X_N \sim \overline{X}(P)$. By Lemma 11,

$$S(X_n) = S(X_{n-1}) + P^n S(\overline{X}(p^n)) \quad \text{for } n = 1, \dots, N.$$

Thus,

$$\begin{aligned}
S(X_N) &= S(X) + \sum_{n=1}^N P^n S(\bar{X}(p^n)) \\
\implies S(X) &= S(X_N) - \sum_{n=1}^N P^n S(\bar{X}(p^n)) \\
&= S(\bar{X}(P)) - \sum_{n=1}^N P^n S(\bar{X}(p^n)) \\
&= \sum_{g=1}^G P_g \log_2 \frac{1}{P_g} - \sum_{n=1}^N P^n \sum_{g=1}^G P_g^n \log_2 \frac{1}{P_g^n}.
\end{aligned}$$

where the last line follows from Claim 7. Q.E.D.

Proof of Proposition 1: IND1: Let X and Y have the same size and group distribution, and let Z be another district. Then $c(X) = c(Y)$ and $T(X)/T(X \uplus Z) = T(Y)/T(Y \uplus Z) = p$. Then, applying SSD, $M(X \uplus Z) \geq M(Y \uplus Z)$ if and only if

$$\begin{aligned}
M(c(X) \uplus c(Z)) + pM(X) + (1-p)M(Z) &\geq M(c(Y) \uplus c(Z)) + pM(Y) + (1-p)M(Z) \\
&\Leftrightarrow M(X) \geq M(Y)
\end{aligned}$$

IND2: Let $W, X, Y \in \mathcal{C}$ be three districts such that $T(W) = T(X)$. Then, $T(W)/T(W \uplus Y) = T(X)/T(X \uplus Y) = p$. Now, applying SSD,

$$\begin{aligned}
M(W \uplus c(Y)) \geq M(X \uplus c(Y)) &\Leftrightarrow M(c(W) \uplus c(Y)) + pM(W) \geq M(c(X) \uplus c(Y)) + pM(X) \\
&\Leftrightarrow M(c(W) \uplus c(Y)) + pM(W) + (1-p)M(Y) \geq M(c(X) \uplus c(Y)) + pM(X) + (1-p)M(Y) \\
&\Leftrightarrow M(W \uplus Y) \geq M(X \uplus Y)
\end{aligned}$$

The proof of GDP is similar and is left to the reader. Q.E.D.

Proof of Proposition 2: Let $X = X^1 \uplus \dots \uplus X^K$ be district composed of K clusters. By

definition of M ,

$$M(X) = h(P(X)) - \sum_{k=1}^K \sum_{n \in \mathbf{N}(X^k)} P^n h(p^n)$$

Subtracting and adding $\sum_{k=1}^K P^k h(P(X^k))$ on the right hand side, we obtain

$$\begin{aligned} M(X) &= h(P(X)) - \sum_{k=1}^K P^k h(P(X^k)) + \sum_{k=1}^K P^k h(P(X^k)) - \sum_{k=1}^K \sum_{n \in \mathbf{N}(X^k)} P^n h(p^n) \\ &= h(P(X)) - \sum_{k=1}^K P^k h(P(X^k)) + \sum_{k=1}^K P^k \left(h(P(X^k)) - \sum_{n \in \mathbf{N}(X^k)} P^n h(p^n) \right) \\ &= M(c(X^1) \uplus \dots \uplus c(X^K)) + \sum_{k=1}^K P^k M(X^k). \end{aligned}$$

This shows that M satisfies SSD. That M satisfies SGD as well now follows from Observation

1. Q.E.D.

References

- [1] Atkinson, A.B. 1970. “On the Measurement of Inequality.” *Journal of Economic Theory* 2: 244-63.
- [2] Bell, Wendell. 1954. “A Probability Model for the Measurement of Ecological Segregation.” *Social Forces* 32:357-364.
- [3] Boozer, Michael A., Alan B. Krueger, and Shari Wolkin. 1992. “Race and School Quality Since Brown v. Board of Education.” *Brookings Papers on Economic Activity: Microeconomics 1992*, pp. 269-338.
- [4] Bourguignon, F. 1979. “Decomposable Income Inequality Measures.” *Econometrica* 47, 901-920.
- [5] Card, David, and Jesse Rothstein. 2006. “Racial Segregation and the Black-White Test Score Gap.” NBER Working Paper No. 12078.

- [6] Carlson, Susan M. 1992. "Trends in Race/Sex Occupational Inequality: Conceptual and Measurement Issues." *Social Problems* 39:269-290.
- [7] Clotfelter, Charles T. 1979. "Alternative Measures of School Desegregation: A Methodological Note." *Land Economics* 54: 373-380.
- [8] Clotfelter, Charles T. 1999. "Public School Segregation in Metropolitan Areas." *Land Economics* 75:487-504.
- [9] Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2002. "Segregation and Resegregation in North Carolina's Public School Classrooms." Terry Sanford Institute of Public Policy working paper no. SAN02-03, Duke University.
- [10] Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. "Achievement and Segregation in Secondary Schools: A Further Look at Public and Private School Differences." *Sociology of Education* 55:162-182.
- [11] Coleman, James S., Sara D. Kelly, and John Moore. 1975. *Trends in School Segregation, 1968-1973*. Washington, D.C.: Urban Institute.
- [12] Collins, Chiquita, and David R. Williams. 1999. "Segregation and Mortality: The Deadly Effects of Racism." *Sociological Forum* 74: 495-523.
- [13] Cortese, Charles F., R. Frank Falk, and Jack K. Cohen. 1976. "Further Considerations on the Methodological Analysis of Segregation Indices." *American Sociological Review* 41:630-637.
- [14] Cotter, David A., JoAnn DeFiore, Joan M. Hermsen, Brenda M. Kowalewski, and Reeve Vanneman. 1997. "All Women Benefit: The Macro-Level Effect of Occupational Integration on Gender Earnings Equality." *American Sociological Review* 62:714-734.
- [15] Cover, Thomas M. and Joy A. Thomas. (1991) *Elements of Information Theory*. United States: Wiley.

- [16] Cutler, David, and Edward Glaeser. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics* 112: 827– 872.
- [17] Duncan, Otis D., and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indices." *American Sociological Review* 20:210-217.
- [18] Echenique, Federico, and Roland G. Fryer, Jr. 2007. "A Measure of Segregation Based on Social Interactions." *Quarterly Journal of Economics* (forthcoming).
- [19] Faddeev, D.K. 1956. "On the Notion of Entropy of a Finite Probability Schema [in Russian]." *Usp. Mat. Nauk* 11:227-231.
- [20] Flückiger, Yves and Jacques Silber 1999. "The Measurement of Segregation in the Labor Force." Physica-Verlag. Heidelberg New York.
- [21] Frankel, David, and Oscar Volij, 2007. "Measuring Segregation: Proofs of Miscellaneous Claims." http://www.econ.iastate.edu/faculty/frankel/segindex_resub_app.pdf.
- [22] Foster, James E. 1983. "An Axiomatic Characterization of the Theil Measure of Income Inequality." *Journal of Economic Theory* 65, 105-121.
- [23] Fuchs, Victor. 1975. "A Note on Sex Segregation in Professional Occupations." *Explorations in Economic Research* 2:105-111.
- [24] Goodman, Leo, and William H. Kruskal. 1954. "Measures of Association for Cross-Classifications." *Journal of the American Statistical Association* 49:732-764.
- [25] Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." Mimeo.
- [26] Herranz, Neus, Ricardo Mora, and Javier Ruiz-Castillo. 2005. "An Algorithm to Reduce the Occupational Space in Gender Segregation Studies." *Journal of Applied Econometrics* 20:25-37.

- [27] Hoxby, Caroline M. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper 7867.
- [28] Hutchens, Robert. 2001. "Numerical Measures of Segregation: Desirable Properties and Their Implications." *Mathematical Social Sciences* 42:13–29.
- [29] Hutchens, Robert. 2003. "One Measure of Segregation." *International Economic Review* 45:555–578.
- [30] Jahn, Julius, and Schmid, Calvin F., and Schrag, Clarence. 1947. "The Measurement of Ecological Segregation." *American Sociological Review* 12:293–303.
- [31] James, Franklin J. 1986. "A New Generalized 'Exposure-based' Segregation Index." *Sociological Methods and Research* 14:301-316.
- [32] James, David R., and Taueber, Karl E. 1985. "Measures of Segregation." In *Sociological Methodology* Jossey-Bass. Nancy Tuma Ed.
- [33] Kinder, Donald R., and Tali Mendelberg. 1995. "Cracks in American Apartheid: The Political Impact of Prejudice among Desegregated Whites." *Journal of Politics* 57:402–424.
- [34] Lewis, Gregory B. 1996. "Gender Integration of Occupations in the Federal Civil Service: Extent and Effects on Male-Female Earnings." *Industrial and Labor Relations Review* 49:472–483.
- [35] Lieberson, Stanley. 1969. "Measuring Population Diversity." *American Sociological Review* 34:850-862.
- [36] Macpherson, David A., and Barry T. Hirsh, 1995. "Wages and Gender Composition: Why do Women's jobs Pay Less? " *Journal of Labor Economics* 13:426–471.
- [37] Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Racial Segregation." *Social Forces* 67:281–315.

- [38] Mora, Ricardo, and Javier Ruiz-Castillo. 2003. "Additively Decomposable Segregation Indices: The Case of Gender Segregation by Occupations and Human Capital Levels in Spain." *Journal of Economic Inequality* 1: 147–179.
- [39] Mora, Ricardo, and Javier Ruiz-Castillo. 2005. "The Axiomatic Properties of an Entropy Based Index of Segregation." Universidad Carlos III de Madrid working paper no. 05-62.
- [40] Mora, Ricardo, and Javier Ruiz-Castillo. 2005. "An Evaluation of an Entropy Based Index of Segregation." Mimeo, Universidad Carlos III de Madrid.
- [41] Mora, Ricardo, and Javier Ruiz-Castillo. 2004. "Gender Segregation by Occupations in the Public and the Private Sector: The Case of Spain." *Investigaciones Económicas* 28:399-428.
- [42] Morgan, Barrie S. 1975. "The Segregation of Socioeconomic Groups in Urban Areas: A Comparative Analysis." *Urban Studies* 12:47-60.
- [43] Philipson, Tomas. 1993. "Social Welfare and Measurement of Segregation." *Journal of Economic Theory* 60:322-334.
- [44] Reardon, Sean F. 1998. "Measures of Racial Diversity and Segregation in Multigroup and Hierarchically Structured Populations." Presented at the Annual Meeting of the Eastern Sociological Society, Philadelphia, PA.
- [45] Reardon, Sean F., and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32:33-67.
- [46] Reardon, Sean F., and John T. Yun. 2001. "Suburban Racial Change and Suburban School Segregation, 1987-95." *Sociology of Education* 74:79-101.
- [47] Sakoda. J. 1981. "A Generalized Index of Dissimilarity." *Demography* 18:245-50.

- [48] Shorrocks, A. F. 1980. "The Class of Additively Decomposable Inequality Measures." *Econometrica* 39:613-25.
- [49] Shorrocks, A. F. 1984. "Inequality Decompositions by Population Subgroups." *Econometrica* 48:1369-86.
- [50] Taeuber, Karl E., and David R. James. 1982. "Racial Segregation Among Public and Private Schools." *Sociology of Education* 55:133-143.
- [51] Taeuber, Karl E., and Alma F. Taeuber. 1976. "A Practitioner's Perspective on the Index of Dissimilarity." *American Sociological Review* 41:884-889
- [52] Theil, Henri. 1971. "Principles of Econometrics." Wiley & Sons. New York.
- [53] Theil, Henri. 1972. "Statistical Decomposition Analysis." North Holland.
- [54] Theil, Henri., and A. J. Finizza 1971. "A Note on the Measurement of Racial Integration in Schools." *Journal of Mathematical Sociology* 1:187-1193.
- [55] Watts, Martin. 1997. "Multidimensional Indexes of Occupational Segregation: A Critical Assessment." *Evaluation Review* 21:461-482.
- [56] Zoloth, Barbara S. 1976. "Alternative Measures of School Segregation." *Land Economics* 52:278-298.