Descriptive Units of Heterogeneity: An Axiomatic Approach to Measuring Heterogeneity

(Latest Version)

Willy Chen*

June 6, 2024†

Abstract

I address the challenge of measuring heterogeneity in a system, as existing measures are unsatisfactory in providing cardinal interpretation and comparability across systems. Using an axiomatic approach, I highlight the strengths and limitations of existing measures and generalize properties that need to be satisfied by alternatives. Using these axioms, I propose a class of measures which I term the *Descriptive Units of Heterogeneity* (DUH), a solution to prior limitations without limiting the applicable contexts. DUH achieves the generalized comparability of concentration units while still being able to reflect changes in the distribution of small groups in the population. I provide several empirical examples demonstrating that DUH is a valuable tool for researchers studying heterogeneity in systems in various contexts, such as racial composition in a city, revenue shares by products of a firm, and trade flows between countries.

Keywords: Diversity, Concentration, Heterogeneity
JEL: B41, D30, D63, J15, L11

*Michigan State University: willyc@msu.edu

†I thank Hanzhe Zhang, Jon Eguia, and Ce Liu for suggestions and feedback.

1 Introduction

I address the challenge of measuring heterogeneity in a system, as existing measures are unsatisfactory in providing cardinal interpretation and comparability across systems (Nunes et al. 2020; Kvålseth 2022). My objective is to measure the degree of heterogeneity present in a system. Take Apple's revenue stream as a system. Heterogeneity in a system means that there is variation in the outcome of interest, and homogeneity is the lack of variation. For example, all else equal, if iPhone sales and iPad sales are 85% and 5% of Apple's revenue, respectively, then the revenue streams of Apple are less heterogeneous than if iPhone sales and iPad sales were each 45% of Apple's revenue. Such examples are easy to understand, but how would we rank two systems with $(55\%, 35\%, 10\%)$ and $(60\%, 20\%, 20\%)$?

Measuring heterogeneity is, at its heart, a dimension-reduction problem. When studying a complex population, one hopes to simplify the complexity without losing sight of the big picture. Choosing the essential elements to *describe* a system is the key to defining a measure that can tractably identify changes in a system. I pin down perfect heterogeneity—as when all groups in the system have the same number of elements and perfect homogeneity—as when all but one group in the system have zero elements. The next task is to (1) balance the influence of large groups and the influence of the distribution of elements across groups and (2) define the role of the number of groups in determining heterogeneity.[1]

The heterogeneity of a system is commonly measured in one of two ways: dispersion units and concentration units (James and Taeuber 1985). Dispersion units measure the distance between the observed population distribution and a benchmark distribution and yield concise interpretation at the cost of comparability between some systems. Atkinson (1970) showed that comparing any two systems using the distance between distributions requires significant restrictions on the domain of comparable systems; otherwise, different measures can be made to rank any two systems in opposing ways. Within the bounds of such restrictions, the interpretation of any dispersion unit is simply "the higher the number, the greater the heterogeneity." One well-known example of a dispersion unit is the *Gini coefficient*. While its simple interpretation contributes to its popularity, Schwartz and Winship (1980) point out that many empirical researchers fail to account for these restrictions when using the Gini coefficient to rank income inequality (pp. 2, 8) and such failure can lead to obscured inferences (pp. 9-13).

Concentration units measure the richness of information from select subgroups of the population. These units differ from dispersion units because of the generalized compatibility between systems, but they understate the information provided by small groups. Both the *Herfindahl-Hirschman index* (HHI) and *Shannon's entropy* (SE) are popular examples of concentration units. By emphasizing the influence of large groups, concentration units well reflect changes in heterogeneity in systems when a large group shrinks. However, this feature can cause concentration units to report negligible changes in heterogeneity when there are drastic changes between small groups. Also, because the objective of these units is to measure concentration, they omit any information provided by the presence of zero-groups—i.e., groups with zero elements. These features are not desirable for a measure of

[1]For example, it may be intuitive to say $\left(\frac{1}{2}, \frac{1}{2}\right)$ is less heterogeneous than $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, but formal reasoning needs to be defined for such inference.

heterogeneity.

Learning from the strengths and weaknesses of these units, I aim to create an index that (1) yields cardinal interpretation, (2) shares basic properties with existing indices, and (3) accounts for the presence of zero-groups. In this paper, I propose a new yet intuitive way to think about the makeup of heterogeneity by separating it into *contribution from the relative size of the largest group* and *contribution from the evenness of the rest of the groups* (minority) and show that this approach builds on the existing paradigm. I divide the desirable properties of measures of heterogeneity into two groups—the *fundamental axioms* and the *characterization axioms*.

Fundamental axioms are axioms that induce partial ordering. They are *Group Symmetry* (SYM), *Scale Invariance* (INV), and the *Principle of Diminishing Transfers* (PDT). The first two axioms pin down that two systems of the same number of groups are equally heterogeneous if one is a permutation or rescaling of the other. PDT pins down how two marginally different systems should be ordered.

Characterization axioms induce total ordering and uniquely characterize measures. I define *Independence* (IND), the *Principle of Proportional Transfers* (PPT), and *Contractibility* (CON). IND ensures that the contribution from the largest group and the evenness of the minority groups are orthogonal; PPT refines PDT and enables my index to yield cardinal interpretation; CON ensures that adding groups with zero elements (zero-groups) decreases heterogeneity. My new axioms, along with the fundamental axioms, characterize a class of indices that focus on making the comparisons between systems *descriptive*—generally comparable, cardinally interpretable, and reflective of changes to small groups. This class of indices is termed the *Descriptive Units of Heterogeneity* (DUH).

Along with the axioms, I put forth the idea of a *reasonable universe of groups*—the set of grouping labels that the researcher deems reasonable and comparable—for practical uses of any measure of heterogeneity. To demonstrate the use cases of DUH, I employ three sets of examples. Using changes in racial heterogeneity in San Francisco from 1900 to 1990, I point out the case where a measure's ability to reflect changes in the evenness in minority groups is pertinent, and how DUH satisfies this need better than HHI and SE. Using changes in Apple's revenue source from 2012 to 2023, I show that DUH is sensitive to changes in the distribution of minority groups and can reflect the growth of a small group better than existing measures. Using changes in international trade flow, I show that concentration units and DUH are complementary in their strengths and should be used optimally with different objectives in mind.

The rest of the paper proceeds as follows. Section 2 discusses existing measures such as the Gini coefficient, HHI, and SE and motivates the need for a new measure. Section 3 lists the fundamental axioms and the characterization axioms along with my proposed new paradigm. Section 4 defines the descriptive units of heterogeneity and compares its behavior to that of the existing concentration units. Section 5 elucidates best practices for application and presents empirical examples to highlight the strength of the descriptive units of heterogeneity in various settings. The appendix provides the requisite proofs that establish the uniqueness of my index under the proposed axioms.

2 Related Literature

This paper is not the first attempt in this strand of literature at an axiomatic characterization of a measure. Rothschild and Stiglitz (1971, 1973) and Schwartz and Winship (1980) axiomatized the *Gini coefficient*, Chakravarty and Eichhorn (1991) and Kvålseth (2022) the *Herfindahl–Hirschman index* (HHI), and Nambiar et al. (1992), Suyari (2004), and Chakrabarti et al. (2005) *Shannon's entropy* (SE). While the Gini coefficient is the simplest form of a unit of heterogeneity in that it operates with the fewest prior assumptions, HHI and SE are more versatile and they are uniquely characterized by their own sets of characterization axioms.

Let $\Theta = \{\theta_1, \dots, \theta_G\}$ be a *universe* of $G \in \mathbb{N}$ distinct groups/categories. A system S is a mapping from Θ to $\mathbb{Z}_+^G$ such that $S = (n_1, \dots, n_g \dots, n_G)$ is a $1 \times G$ vector where n_g is a positive integer that represents the number of elements in the group θ_g. A system S with population n_S is thus the collection of groups θ_g each with n_g elements.

The measure of heterogeneity is then a mapping $\Phi : \mathbb{Z}_+^G \to \mathbb{R}$ such that for any two systems S and S',

$$\Phi(S) \geq \Phi(S') \iff S \text{ is weakly more heterogeneous than } S'.$$

For example, consider the population of Michigan State University a system S. S maps the universe of groups $\Theta = \{faculty, \; staff, \; students\}$ to the number of faculty (n_{fac}), staff (n_{sta}), and students (n_{stu}) at Michigan State. Heterogeneity in this system is the presence of mixture, e.g. the presence of a mix of faculty, staff, and students. Homogeneity is the lack of mixture, meaning only one or two of these groups are present in the system. I can thus define mathematically what it means for a system to be maximally/minimally heterogeneous.

Definition 1: A system S_{max} of G groups is said to achieve **maximum heterogeneity** if it can be represented as a scalar multiple of the identity vector of size $G \in \mathbb{N}$:

$$S_{max} = (\underbrace{n, n, \dots, n}_{\substack{G \text{ groups each} \\ \text{with } n \text{ elements}}}) = n \cdot (1, 1, \dots, 1).$$

Definition 2: A system S_{min} of G groups is said to achieve **minimum heterogeneity/perfect homogeneity** if it can be represented as a $1 \times G$ vector where all but one entry are 0:

$$S_{min} = (0, 0, \dots, 0, n, 0, \dots, 0) = n \cdot (0, 0, \dots, 0, 1, 0, \dots, 0).$$

The one-dimensional (presence of mixture) nature of this definition makes it convenient for any measure to be bounded between $\Phi(S_{min}) = 0$ and $\Phi(S_{max}) \in \mathbb{R}_{++}$.[2] These units can be generally separated into two categories—dispersion units and concentration units (James and Taeuber 1985).

Dispersion units.—This class of units compare the observed distribution to a benchmark distribution. The most commonly used dispersion unit is the Gini coefficient. It compares the

[2]Typical measures of heterogeneity are bounded between 0 and 1.

distribution of an outcome to the uniform distribution and measures heterogeneity based on "the top x% of households in the US earns the top y% of income." In a uniformly distributed world (maximum heterogeneity), x and y should be equal, and if it is not, then there is inequality. This simple interpretation comes at the cost of a restrictive assumption—the *Lorenz criterion*—where the order of the Gini coefficients of two systems infers the order of heterogeneity between them if and only if their Lorenz curves do not cross (Atkinson 1970; Marshall et al. 1979; Schwartz and Winship 1980; James and Taeuber 1985).[3] However, the common limitations of dispersion units make them less practical in modern empirical research where one may need a general comparison between systems.

Concentration Units.—Both HHI and SE are derived from the *Hannah-Kay* class of concentration units (Hannah and Kay 1977), with perception α and $G \in \mathbb{N}$ firms in industry $S \in D^G$, defined as:

$$H_\alpha^G(S) = \begin{cases} \left[\sum_{g=1}^{G} P_g^\alpha\right]^{\frac{1}{\alpha-1}} & \text{if } \alpha > 0,\ \alpha \neq 1 \\ \prod_{g=1}^{G} P_g^{P_g} & \text{if } \alpha = 1 \end{cases}, \quad P_g = \frac{n_g}{n_S}.$$

Consider a system S with G groups. HHI and its complement *Gini-Simpson index* (GSI) of system S are defined as:

$$HHI(S) = \sum_{g=1}^{G} \left(\frac{n_g}{n_S}\right)^2 = H_2^G(S), \quad GSI(S) = 1 - HHI(S).$$

SE is defined as:

$$SE(S) = -\sum_{g=1}^{G} \left[\frac{n_g}{n_S} \cdot ln\left(\frac{n_g}{n_S}\right)\right] = -ln\left(H_1^G(S)\right).$$

HHI is comparable across systems and it yields cardinal interpretation—the probability of 2 random draws with replacement, from the system S, being from the same group. Nevertheless, it disproportionately accounts for changes in large groups, which is desirable in assessing the market power of firms (Chakravarty and Eichhorn 1991) or the power of political parties (Laakso and Taagepera 1979). Consequently, the significance of an 80% reduction in a firm's presence, initially constituting a mere 5% of the market, may appear negligible if the 4% were redistributed to another minor competitor. This necessitates contextual confinements of HHI's application to discussions concerning market shares and concentrations (Kvålseth 2022), rather than employing it as a measure of heterogeneity.[4]

SE is a popular measure of uncertainty/informativeness in the information theory and rational inattention literature (Sims 2010, 2003; Pomatto et al. 2023). Notice that SE is

[3] Nunes et al. (2020) discusses several other dispersion units in the context of ecology, biology, and medicine.

[4] As another example, the simple interpretation of HHI can obscure obvious differences between systems; two systems, $(48\%, 48\%, 4\%)$ and $(60\%, 30\%, 10\%)$ will yield 0.4624 and 0.46 HHI. The heterogeneity of these systems is reasonably different, but that is hardly reflected in the difference in their HHIs. In comparison, the DUHs of these two systems are 0.2684 and 0.3006, indicating the second system is much more heterogeneous.

a negative logarithmic transformation of H_1^G and hence it (1) loses the probabilistic interpretation of HHI and (2) behaves differently when groups are broken up. One key thing to notice is that in both of these units, zero-groups—i.e., groups with zero elements—do not affect the measure at all. This property is intuitive when used to capture market shares or uncertainty, but it should not be salient when used to measure heterogeneity.

Comparability between systems hinges on what the comparison is to capture and whether the two systems are similar enough.[5] To effectively compare heterogeneity between two systems, it is crucial to establish the qualifiers that make them comparable. Including information from all groups, even those with zero elements, provides a baseline for measurement. Thus, comparing systems with varying group counts requires viewing the system with fewer groups as encompassing the additional groups with zero elements, ensuring an accurate evaluation of heterogeneity.

Catering an index to the inclusion of zero-groups is hardly revolutionary. A normalized version of HHI attempts to solve that issue by revising the formula to:

$$\mathit{NHHI}(S,G) = \frac{HHI(S) - \frac{1}{G}}{1 - \frac{1}{G}} \in [0,1].$$

This index improves system comparability by accounting for zero-groups via normalization, but it is done at the cost of HHI's probabilistic interpretation.[6]

3 Axioms for Units of Heterogeneity

Fundamental Axioms.—First, I start with the set of axioms that induce partial ordering. These axioms pin down when two systems are equally heterogeneous and how two systems should be ordered when one is a simple elementary transfer of the other.

[SYM] Group Symmetry. For any permutation $\pi(S)$ of S, $\Phi(S) = \Phi(\pi(S))$.

For example, take $n_a, n_b, n_c \in \mathbb{N}$,

[5] Comparing apples to oranges might not necessarily be nonsensical as the popular idiom suggests. In the proper context, I can compare an apple to an orange in its density, brightness of color, amount of sugar per milliliter of water, etc. I can even compare an apple to an orange on how good each fruit is at being citrus (clearly, the apple will lose). What I cannot do is say that an apple is x times denser than an ice cube and an orange is y times rounder than a bowling ball.

[6] Consider the following two systems:

$$S = (0.4, 0.4, 0.2) \text{ and } S' = (0.5, 0.3, 0.1, 0.1).$$

These two systems have the same HHI (0.36), but they have different NHHIs ($\mathit{NHHI}(S, G = 3) = 0.04$ and $\mathit{NHHI}(S',\, G = 4) \approx 0.15$). By observation, it may not be clear whether S and S' are equally homogeneous, but the comparison of these two NHHIs is unlikely to be convincing. Once we account for zero-groups and make $S = (0.4, 0.4, 0.2, 0)$, the NHHIs of the two systems are the same (0.15) just like their HHIs, but the level of heterogeneity can no longer be intuitively interpreted.

$$
\begin{array}{|c|}\hline \text{G. \textbackslash\ S.} \\ \hline \text{A} \\ \hline \text{B} \\ \hline \text{C} \\ \hline \end{array}
\quad
\begin{array}{|c|}\hline S_1 \\ \hline n_a \\ \hline n_b \\ \hline n_c \\ \hline \end{array}
\sim
\begin{array}{|c|}\hline S_2 \\ \hline n_b \\ \hline n_a \\ \hline n_c \\ \hline \end{array}
\sim
\begin{array}{|c|}\hline S_3 \\ \hline n_c \\ \hline n_b \\ \hline n_a \\ \hline \end{array}
$$

SYM is an intuitive axiom as it enables the index to focus on the distribution over groups in a system, rather than the sizes of individual groups. Satisfying SYM means any two systems with the same number of groups can be compared. By focusing on distribution over the same number of groups, SYM enables comparisons between systems mapping different universes of groups, so long as the two universes have the same number of groups.

> **[INV] Scale Invariance.** For any system S and a scalar $\lambda \in \mathbb{R}_{++}$, $\Phi(S) = \Phi(\lambda \cdot S)$.

INV allows for a further generalization of the systems to have groups with sizes of any non-negative real number. This axiom ensures that the index reflects only the distribution of sizes in a system rather than the absolute sizes of the groups.

> **[PDT] Principle of Diminishing Transfers.** *Holding the order of groups constant*, a transfer from a larger group to a smaller group increases heterogeneity. The increase increases in the difference between the two groups.

For example, take $n_a, n_b, n_c \in \mathbb{N}$ such that $n_a > n_b > n_c$ and $\varepsilon < \min\left\{n_b - n_c, \dfrac{n_a - n_b}{2}\right\}$.

$$
\begin{array}{|c|}\hline \text{G. \textbackslash\ S.} \\ \hline \text{A} \\ \hline \text{B} \\ \hline \text{C} \\ \hline \end{array}
\quad
\begin{array}{|c|}\hline S_1 \\ \hline n_a \\ \hline n_b \\ \hline n_c \\ \hline \end{array}
\prec
\begin{array}{|c|}\hline S_2 \\ \hline n_a - \varepsilon \\ \hline n_b + \varepsilon \\ \hline n_c \\ \hline \end{array}
\prec
\begin{array}{|c|}\hline S_3 \\ \hline n_a - \varepsilon \\ \hline n_b \\ \hline n_c + \varepsilon \\ \hline \end{array}
$$

This axiom originated as the *Principle of Transfers*, first formulated by Dalton (1920), "...if there are only two income-receivers, and a transfer of income takes place from the richer to the poorer, inequality is diminished" (p.351). For a measure of heterogeneity, I believe that the decrease in inequality should be increasing in the difference of proportions of the two groups, adding the "diminishing" to the axiom formulated by Dalton (1920).

A New Way to Think About Heterogeneity.—Gini, HHI, and SE treat each group identically, ensuring SYM. However, SYM can also be satisfied by treating the same types of groups identically, rather than each group. Notice that for any system, there is always the largest group and the remaining groups. So an index that treats the largest group differently than the rest of the groups in the system can still satisfy SYM. The advantage of this new perspective on heterogeneity is that if the largest group's influence on the index is orthogonal to the rest, changes in the one-dimensional index can be equated to changes in either group while keeping the other constant.

For convenience, I refer to groups that are not the largest group the *minority groups*. Under this new paradigm, an index of heterogeneity is then $\Phi = \Phi(\varphi, \psi)$ where $\varphi(P_1)$ is the influence of the relative sizes of the largest group and $\psi(P_2, \ldots, P_G)$ the influence of the relative sizes of the minority groups.

Characterization Axioms.—My next step is to build on the fundamental axioms to gain total ordering and uniquely characterize my measure. In existing literature, this is done by imposing cardinal interpretation and prescribing how the measure changes when groups are broken up into smaller groups.

[IND] Independence. The influence of the relative size of the largest group should be independent of the relative sizes of the minority groups, and vice versa.

$$\varphi(n_1, n_2, \ldots, n_G) = \varphi\left(\frac{n_1}{n_1 + \cdots + n_G}\right) = \varphi(P_1).$$
$$\psi(n_1, n_2, \ldots, n_G) = \psi(n_2, n_3, \ldots, n_G) = \psi(P_2, \ldots, P_G).$$

IND enables us to determine what an equivalent change in heterogeneity would look like if it resulted solely from changes in either the size of the largest group or the evenness of the minority groups; in reality, it often lies between these extremes. To accommodate an arbitrary number of groups, ψ needs to capture features of the minority group distribution. I define ψ to be a function of the distance between the observed distribution in the minority groups and the ideal uniform distribution in the minority groups.

Definition 3: A function $\psi : \mathbb{R}_+^{G-1} \to \mathbb{R}_+$ is a measure of evenness in minority group distribution if it is of the following form:

$$p \in \mathbb{R}_+,\ \psi_p(S) = \psi\left(\tilde{P}_2, \ldots, \tilde{P}_G\right) = 1 - \left(\sum_{g=2}^{G}\left|\tilde{P}_g - \frac{1}{G-1}\right|^p\right)^{\frac{1}{p}}.$$

Proposition 1: Consider an index $\Phi_p = \Phi(\varphi, \psi)$ that satisfies SYM, INV, and IND. Holding P_1 constant, if Φ is strictly increasing in ψ_p, then $\Phi(\varphi(P_1), \psi)$ satisfies PDT if and only if $p > 1$.[7]

Having formally defined ψ_p, I want to pin down φ with an axiom that fully describes changes in heterogeneity when ψ is fixed. In pursuit of an index whose changes are easy to interpret, I propose a minimalist refinement of PDT that builds on the notion of IND—*Principle of Proportional Transfers.*

[PPT] Principle of Proportional Transfers. *Holding the order of groups constant*, a transfer from the largest group proportionally to the minority groups that reduces P_1 to $(P_1)^\alpha$ increases heterogeneity by a factor of α.

[7]This proposition suggests that we need to be careful with the functional form of $\varphi(P_1)$. To satisfy PDT, it must be that when there is a transfer from the largest group to the minority group, the increase in $\varphi(P_1)$ must dominate the decrease in ψ_p in the case where evenness decreases. Additionally, if $G = 3$, then this proposition holds with $p \geq 1$.

For example,

G. \ S.	S_1		S_2
θ_1	P_1		P_1^{α}
θ_2	P_2	$\prec$	$P_2 + \frac{P_2}{P_2+P_3}(P_1 - P_1^{\alpha})$
θ_3	P_3		$P_3 + \frac{P_3}{P_2+P_3}(P_1 - P_1^{\alpha})$

Then

$$\Phi(S_1) < \alpha\Phi(S_1) = \Phi(S_2).$$

PPT gives changes in the index a simple interpretation. If $2\Phi(S_1) = \Phi(S_2)$, then one can say that S_2 *is twice as diverse as* S_1 because it has the equivalent heterogeneity as if the largest group proportion in S_1 shrunk by the power of 2 while still being the largest group, holding the same evenness in the minority groups.

[CON] Contractibility. Φ satisfies *Contractibility* if adding one 0-group to a system of G groups decreases heterogeneity of the system.

$$\Phi(n_2, \ldots, n_G, 0) < \Phi(n_2, \ldots, n_G).$$

A practical implication of CON is that the comparison between systems with a unit assumes that the two systems have the same number of groups, even if some groups have 0 elements. For example, $\left(\frac{1}{2}, \frac{1}{2}, 0\right) \prec \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Axioms used for the characterization of HHI and SE are presented in appendix B.

4 The Descriptive Units of Heterogeneity

I propose the Descriptive Units of Heterogeneity —a class of units that balance interpretability and comparability.

Definition: Let $n_1 \geq n_2 > 0$, $P_1 = \frac{n_1}{n_1+n_2+\cdots+n_G}$, and $\tilde{P}_g = \frac{n_g}{n_2+\cdots+n_G}$, $g > 1$. The Descriptive Units of Heterogeneity (DUH) of the system S with $G \geq 2$ groups is:

$$DUH(S) = \frac{ln(P_1)}{ln(G)} \cdot \left[\left(\sum_{g=2}^{G} \left| \tilde{P}_g - \frac{1}{G-1} \right|^p \right)^{\frac{1}{p}} - 1 \right].$$

Theorem: The Descriptive Units of Heterogeneity is the unique class of units, up to positive scalar multiplication, that satisfies Scale Invariance, Group Symmetry, Independence, Principle of Diminishing Transfers, Principle of Proportional Transfers, Contractibility, and uses ψ_p to incorporate the measure of evenness.

Proof sketch: (1) I show that any index using an ordered vector of group proportions satisfies SYM and INV. (2) I then show that any $\Phi(\varphi, \psi)$ satisfying SYM, INV, and PPT must be a positive monotonic transformation of $\frac{1}{P_1}$. (3) I show if an $\Phi(\varphi, \psi)$ satisfies INV, PPT, and IND, then $\Phi = \varphi \cdot \psi$. (4) I show that if $\Phi(\varphi, \psi)$ satisfies INV, IND, and PPT, then $\varphi = -c ln(P_1)$, $c \in \mathbb{R}_{++}$. (5) I show that the theorem is true by taking the derivative of the extreme case where P_1 is close to 1 and $\psi = 1$ with respect to a transfer from the largest group to the second largest group.

Notice that the parameter p controls how much evenness is reflected in DUH through ψ_p. As p increases, minority groups that are farther away from $\frac{1}{G-1}$ take more weight. As $p \to \infty$,

$$d_p \to d_\infty = \sup_{g \in \{1,\ldots,G\}} \left\{ \left| P_g - \frac{1}{G-1} \right| \right\}.$$

Figure 1 shows how the progression of DUH changes with different p's. As p increases, the contribution of the evenness in the minority takes less weight. When $p = 2000$, the effects of transfers between minority groups become negligible, making it look like P_1 dominates evenness in calculating DUH.

Figure 1: DUH with Different p

$p = 20$ $\quad$ $p = 200$ $\quad$ $p = 2000$

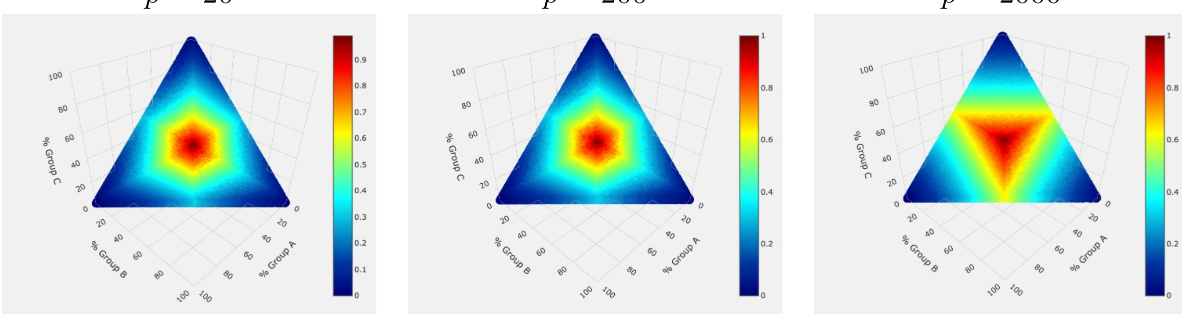


Table 1 outlines the axioms discussed and whether Gini, DUH, HHI, or SE satisfy them. Notice that DUH can be considered a refined Gini coefficient with generalized comparability across systems made up of discrete and unordered groups. DUH builds on the partial ordering of the Gini coefficient to induce a total order that enables better comparisons across systems that do not satisfy the Lorenz criterion by further refining PDT and incorporating ψ_p. DUH is characterized differently than concentration units because it focuses on the overall distribution without losing comparability. Note that since all measures here satisfy the fundamental axioms, if the distributions of two systems satisfy the Lorenz criterion, then the Gini coefficient induces the same order between the two systems as the other three measures.

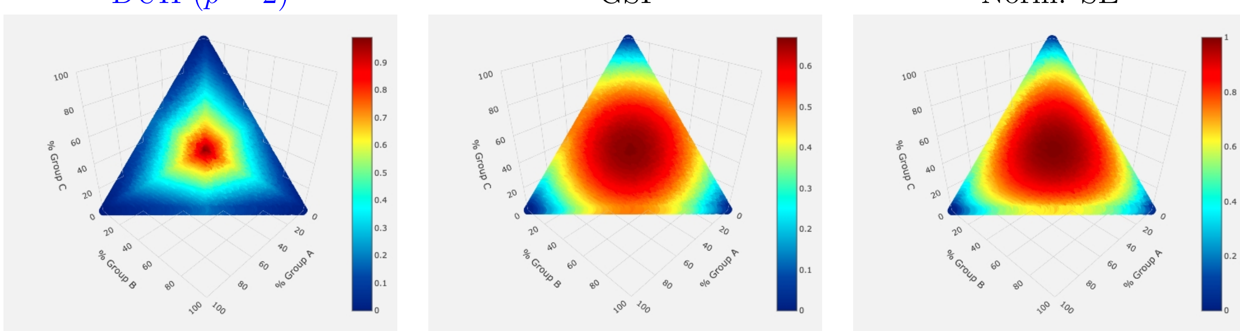
Table 1: Measures and Axioms

Type	Axiom	Gini	DUH	HHI	SE
Fundamental	Type Symmetry	✓	✓	✓	✓
	Scale Invariance	✓	✓	✓	✓
	Principle of Diminishing Transfers	✓	✓	✓	✓
Characterization	Independence	×	✓	✓	✓
	Principle of Proportional Transfers	×	✓	×	×
	Contractibility	✓	✓	×	×
	Expandability[8]	×	×	✓	✓
	Replication Principle[8]	×	×	✓	×
	Shannon's Additivity[8]	×	×	×	✓

Figure 2 compares DUH to the concentration units where $G = 3$. The tetrahedrons of each measure below show how each measure changes as the distribution of groups becomes more heterogeneous.[9] The centers of the triangles represent a perfectly heterogeneous system, and the vertices of the triangles represent perfectly homogeneous systems.

Figure 2: Differences between DUH, GSI, and SE

DUH ($p = 2$) GSI Norm. SE



[8]See appendix B for details of these axioms and characterizations
[9]SE is normalized to be between 0 and 1 by dividing it by $ln(3)$.

5 Practical Uses of DUH

DUH, as a simple description of heterogeneity in systems, can be used in various contexts with discrete distributions over unordered groups in a system. This section provides several empirical examples where the strength of DUH is shown. Since HHI is a measure of concentration/homogeneity between 0 and 1, I use GSI (=1-HHI) here to make the comparisons simpler. Similarly, SE is normalized to be between 0 and 1 to make the comparisons simpler.

5.1 Reasonable Universe Θ

Recall that to measure the heterogeneity of any system, the system must first be thought of as a mapping from a universe of groups. Let $\Theta = \{\theta_1, \ldots, \theta_G\}$ be a **universe** of G distinct groups/categories. A system S is a mapping from Θ to $\mathbb{Z}_+^G$ such that $S = (n_1, \ldots, n_g \ldots, n_G)$ is a $1 \times G$ vector where n_g is a positive integer that represents the number of elements in θ_g in the system S with population n_S.

This paradigm implies that the set Θ needs to be handled with care because each $\theta \in \Theta$ must be similar/comparable to each other. Figures 3 and 4 illustrate this idea with a practical example. These two figures present the racial composition of San Francisco MSA from 2007 to 2022 using ACS 1-year data (Ruggles et al. 2024). Figure 3 defines Θ as {White, Black, Other} while figure 4 splits up the *Other* group into 3 sub-groups, yielding $\Theta =$ {White, Black, Asian, Native America, and Multi-Race}.

In figure 3, the heterogeneity of this system is somewhat stable due to the influence of the shrinkage in the white population and the increase in the other population. The heterogeneity started to decrease post-2019 when the *white* population became a minority group and the *other* population became the largest group. This change shows the importance of SYM which allows researchers to study heterogeneity as a distributional property free of labels. However, the story is different once Θ is redefined to further capture distributional changes in subgroups. Figure 4 shows that when the Asian population and multi-race population are considered separately, heterogeneity increases post-2019, as the groups, at a glance, are proportionally growing. Such distributional changes are what PPT is designed to reflect.

When measuring heterogeneity in a system, one must realize the implications of choosing Θ. Determining the elements of Θ is a framing problem and is a judgment call by the researcher. Just as the use of Gini coefficient requires the Lorenz Criterion, the use of any units of heterogeneity requires justification of the reasonable groupings. In the examples here, the simple split of a subgroup changed the inference, serving as an excellent reason why these units need to be used with much care.

Keeping this in mind, let us consider examples of when DUH can be used and why it should be used. For simplicity, I will use my preferred version of DUH where $p = 2$ so ψ uses the Euclidean distance.

5.2 Examples

The examples here demonstrate how DUH can be useful for interpreting heterogeneity in different environments. The examples are arraigned to progress in the size of the reasonable

Figure 3: Comparisons between Different Units for Racial Heterogeneity

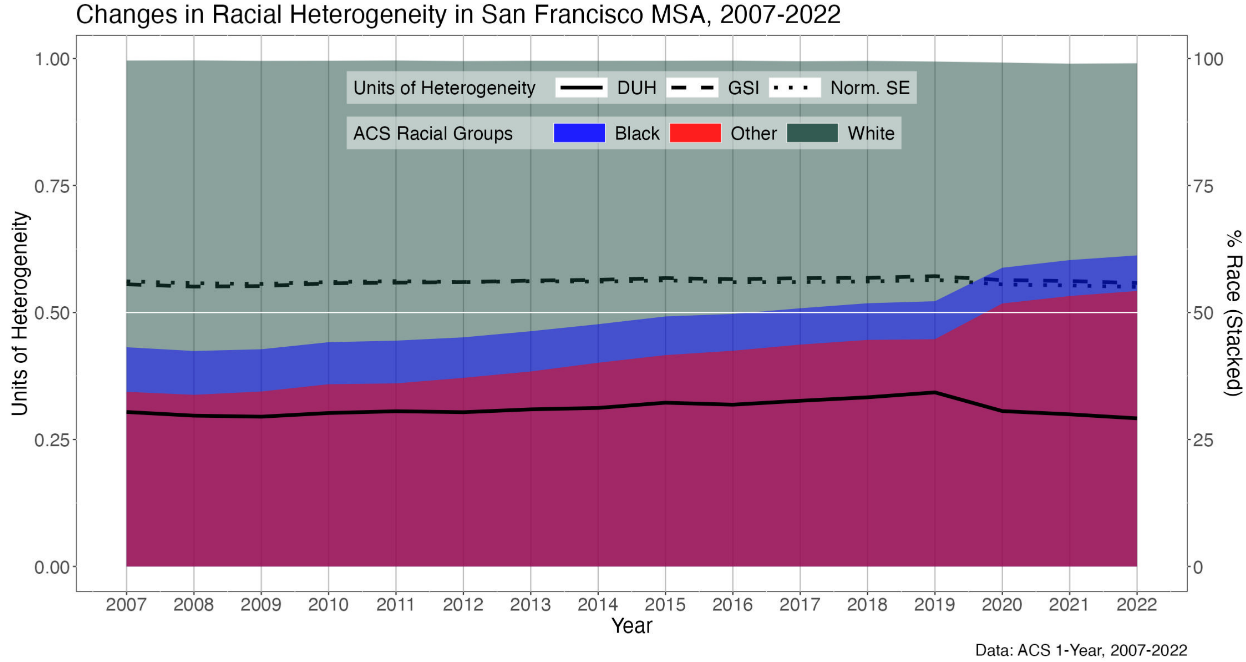
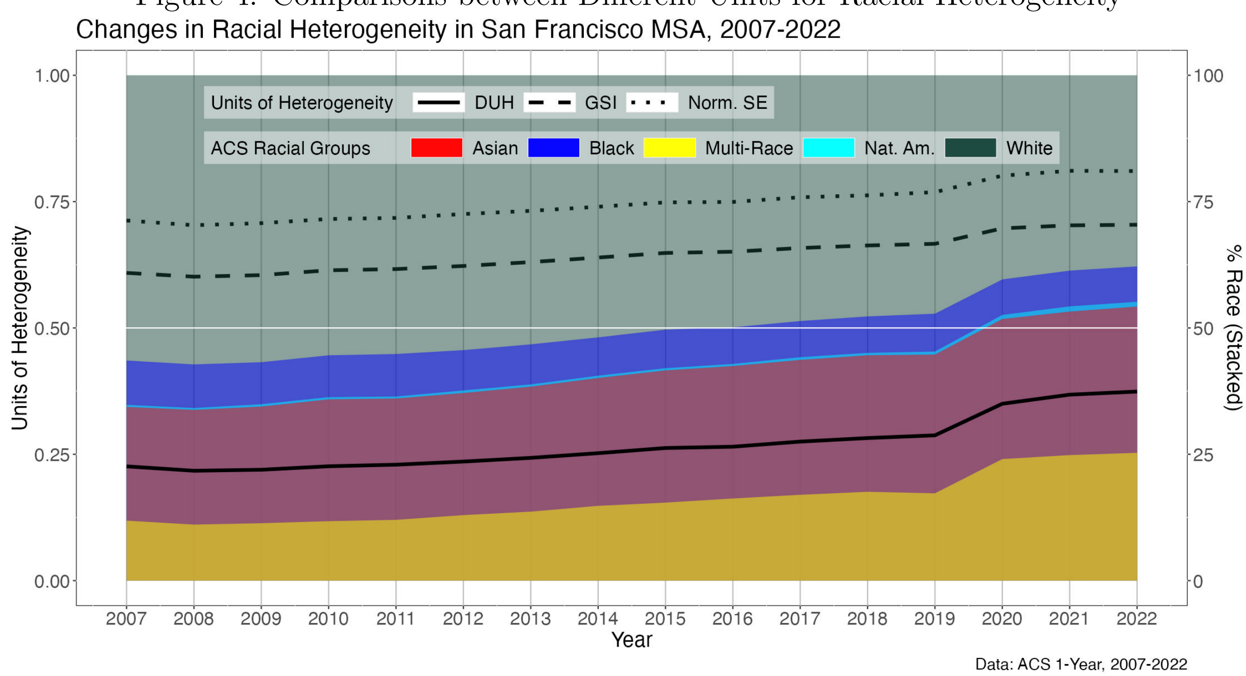


Figure 4: Comparisons between Different Units for Racial Heterogeneity

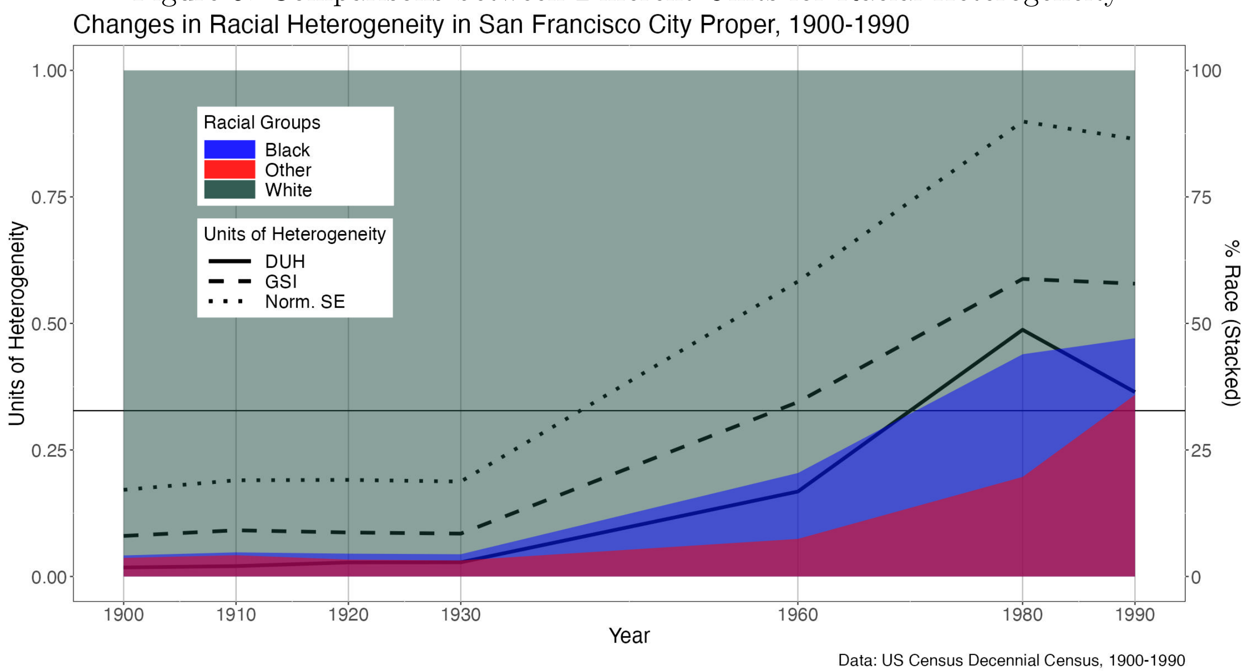


universe of groups to show that DUH is sensitive only to the two components— the size of the largest group and the evenness of minority groups—and not to the number of groups.

5.2.1 Using DUH for Racial Heterogeneity

The first example uses DUH to measure racial heterogeneity when there are only 3 groups—White, Black, and Other—in the reasonable universe Θ. Figure 5 shows the progression of racial heterogeneity in San Francisco city proper from 1900-1990 using the decennial Census data from IPUMS USA (Ruggles et al. 2024).[10]

Figure 5: Comparisons between Different Units for Racial Heterogeneity



From 1980 to 1990, the population of the largest group (white) of San Francisco city proper decreased slightly, but the other population (mostly Asian) grew so much that it made the minority groups distributions much less even. In this case, GSI indicated only a slight decrease in heterogeneity while the larger decrease in SE reflects more of this change in the minority group distribution. DUH, on the other hand, follows generally the same trends as GSI and SE, yet it can reflect much more of the decrease in evenness in the minority distribution.

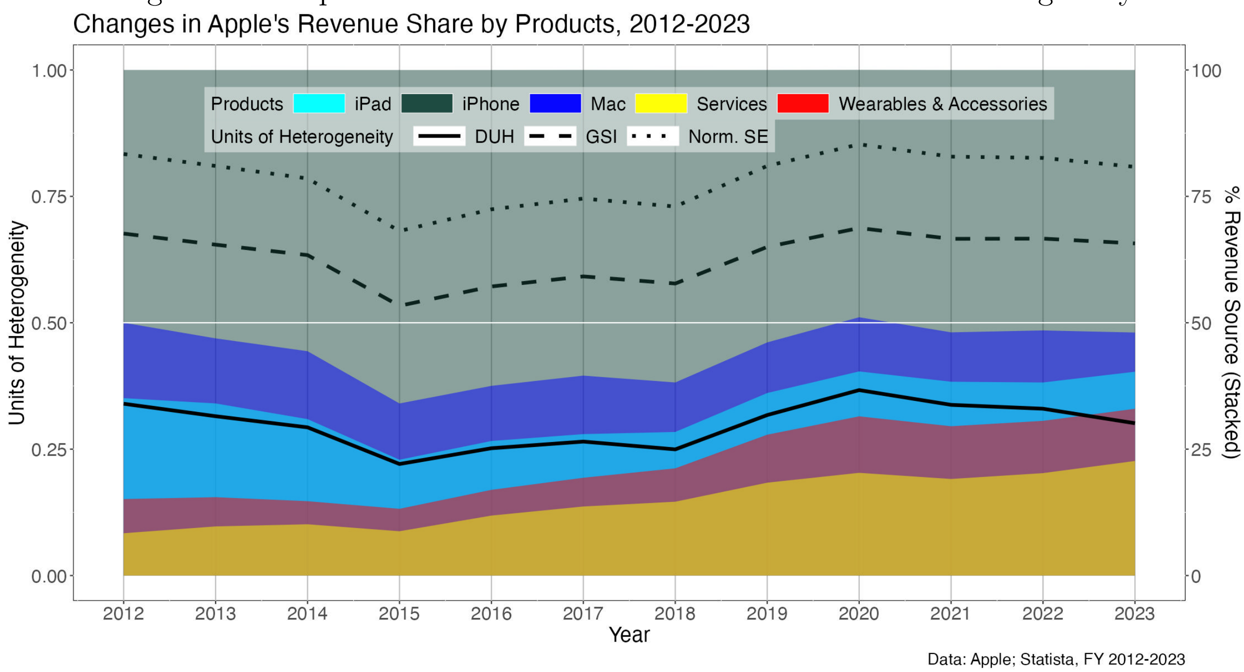
Recall that the main weakness in GSI and SE is that the size of the influence from changes in a group positively correlates with the size of the group. This example shows that DUH can dial back the correlation and reflect changes in the evenness of the minority groups.

[10] Due to Census coding of the inner-city variable, data is missing for 1940, 1950, and 1970.

5.2.2 Using DUH for Revenue Heterogeneity

The second example uses DUH to proxy how lightly a firm relies on specific products for its revenue. In this example, there are 5 groups—iPhone, iPad, Mac, Wearables & Accessories, and Services—in the reasonable universe for Apple's revenue. Figure 6 illustrates how DUH compares with GSI and DUH in a space that often utilizes units of concentration using data on Apple's revenue source by products (Apple and Statista 2024). For the most part, the three units move in the same way. However, notice that from 2020 to 2023, Apple's revenue share for services as well as wearables (like Apple watch) & accessories (like AirPods) grew without diminishing the revenue share of iPhones. This decrease in the evenness in minority groups is captured by a continuous and sizable decrease in DUH, while decreases in GSI in this period are limited.

Figure 6: Comparisons between Different Units for Revenue Heterogeneity



5.2.3 Using DUH for Trade Share Heterogeneity

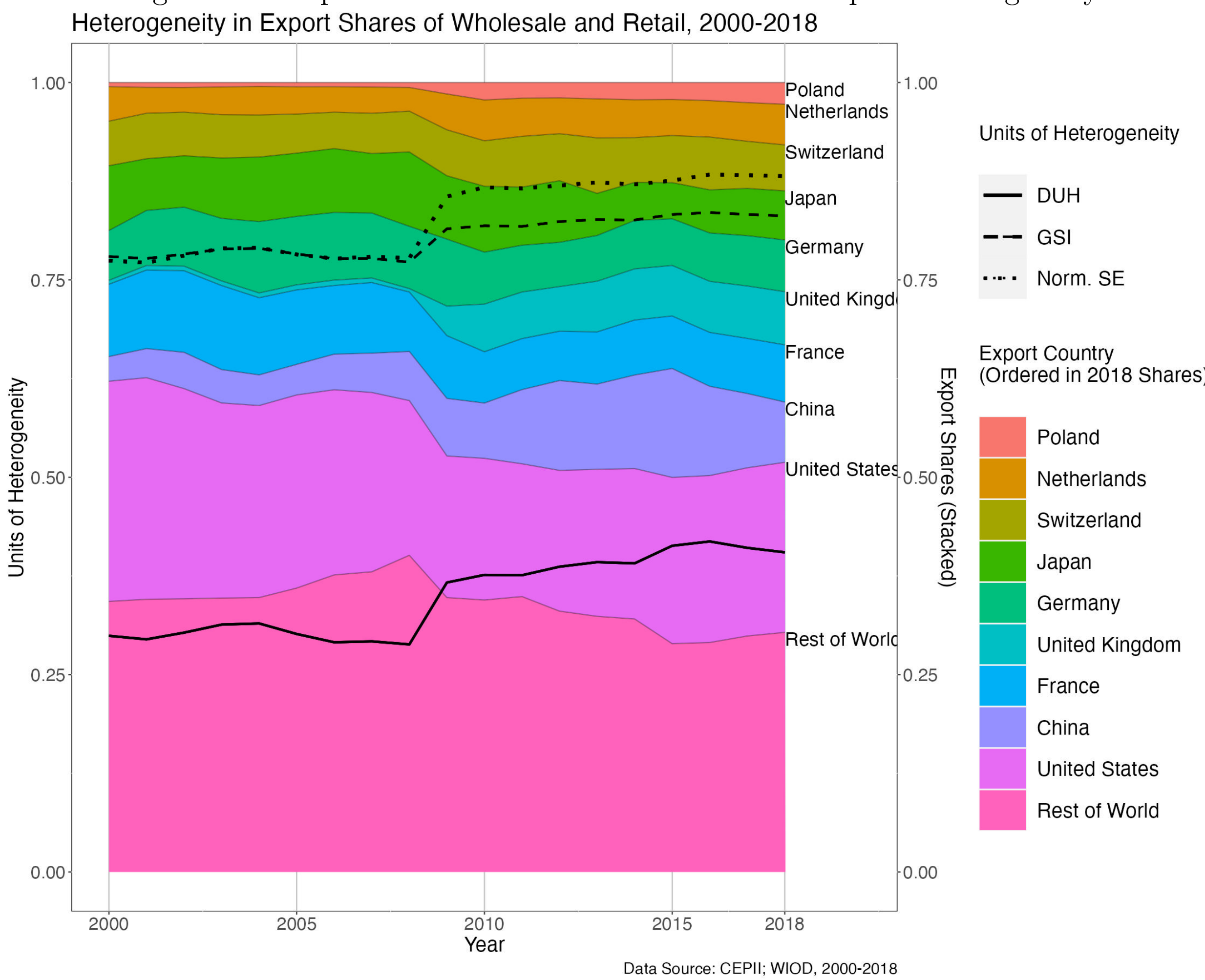
The last set of examples uses DUH in the environment where HHI and SE are typically used—Trade. In this setting, it is reasonable to focus on the concentration and not the heterogeneity in the system. As I progress through the three examples, readers should notice why a measure of heterogeneity does not necessarily need to agree with a measure of concentration. Using trade flow data from CEPII-BACI (Gaulier and Zignago 2010) and the World Input-Output Database (WIOD) (Timmer et al. 2012), I calculated the export share of each of the 29+1 countries,[11] import shares from each of the 28+1 countries to the US,

[11] 29 of 164 member countries of the World Trade Organization as well as the rest of the world as one country.

and export shares to each of the 28+1 countries from the US[12].

Within-Industry Export Shares of Countries.[13]—Figure 7 graphs the global export shares of the top 9+1 countries (in 2018) for *Wholesale and Retail*. In 2009, there was a drastic increase in all three measures, coinciding with both the significant growth in export shares of the United Kingdom and Poland and the shrinkage in the share of the rest of the world (RoW). From 2010 to 2015, the export shares of the US and China grew as the shares of RoW decreased. This growth is well-reflected in DUH but changes in GSI and SE are minimal. Similarly, from 2016 to 2018, the export share of China shrunk as the export shares of RoW grew, resulting in a decrease in DUH while changes in GSI and SE remained minimal.

Figure 7: Comparisons between Different Units for Export Heterogeneity



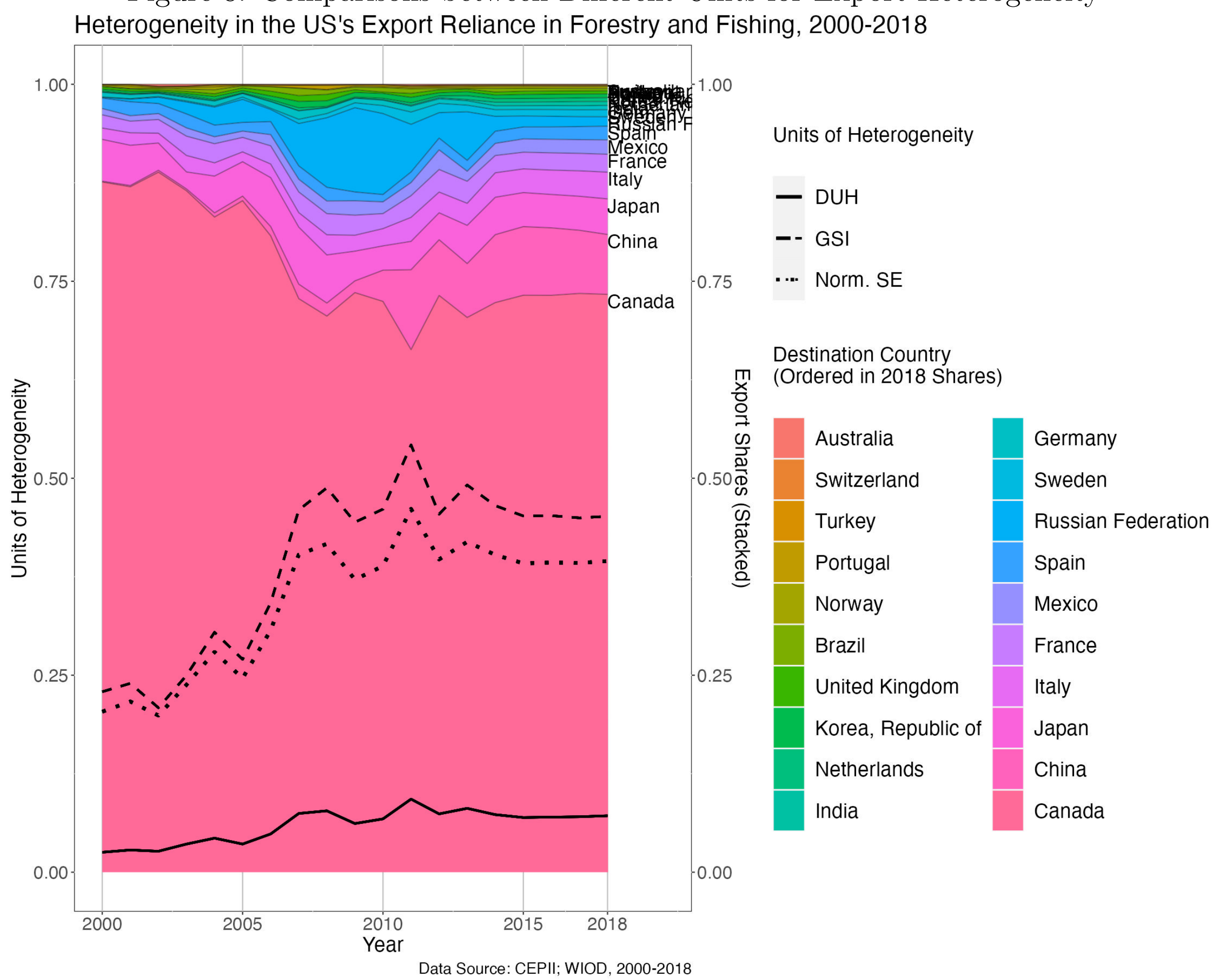
This example shows how a DUH reflects changes in small groups better than HHI and SE. The next two examples demonstrate why DUH is not a substitute of concentration units.

[12]Special thanks to my friend Erin Eidschun, an Economics PhD student at Boston University, who graciously provided a cleaned and organized version of this data set. As these are only illustrative examples, I am less concerned about any inaccuracies there may be.

[13]The share of all exported goods in the industry from the country.

Export Reliance.—Export reliance is similar to the revenue stream. If a country's export in an industry heavily relies on one or a few countries, then one can argue that said country has little power in determining trade policy regarding that industry. Figure 8 graphs the share of the US's exports in *Forestry and Fishing* to the top 20 destination countries.[14] Changes in DUH are relatively minor compared to changes in GSI and SE. Due to PDT, the range of influence of evenness in the minority decreases in the size of the largest group. In figure 8, without looking at the shares directly, DUH suggests that the US's exports of forestry and fishing were heavily reliant on one or a few countries in 2000-2018; Both GSI and SE suggests an initial reliance on a few countries, followed by a decrease in export reliance in 2007. Comparing that to the shares, whether DUH or the concentration units are more descriptive of export reliance is at the researcher's discretion.

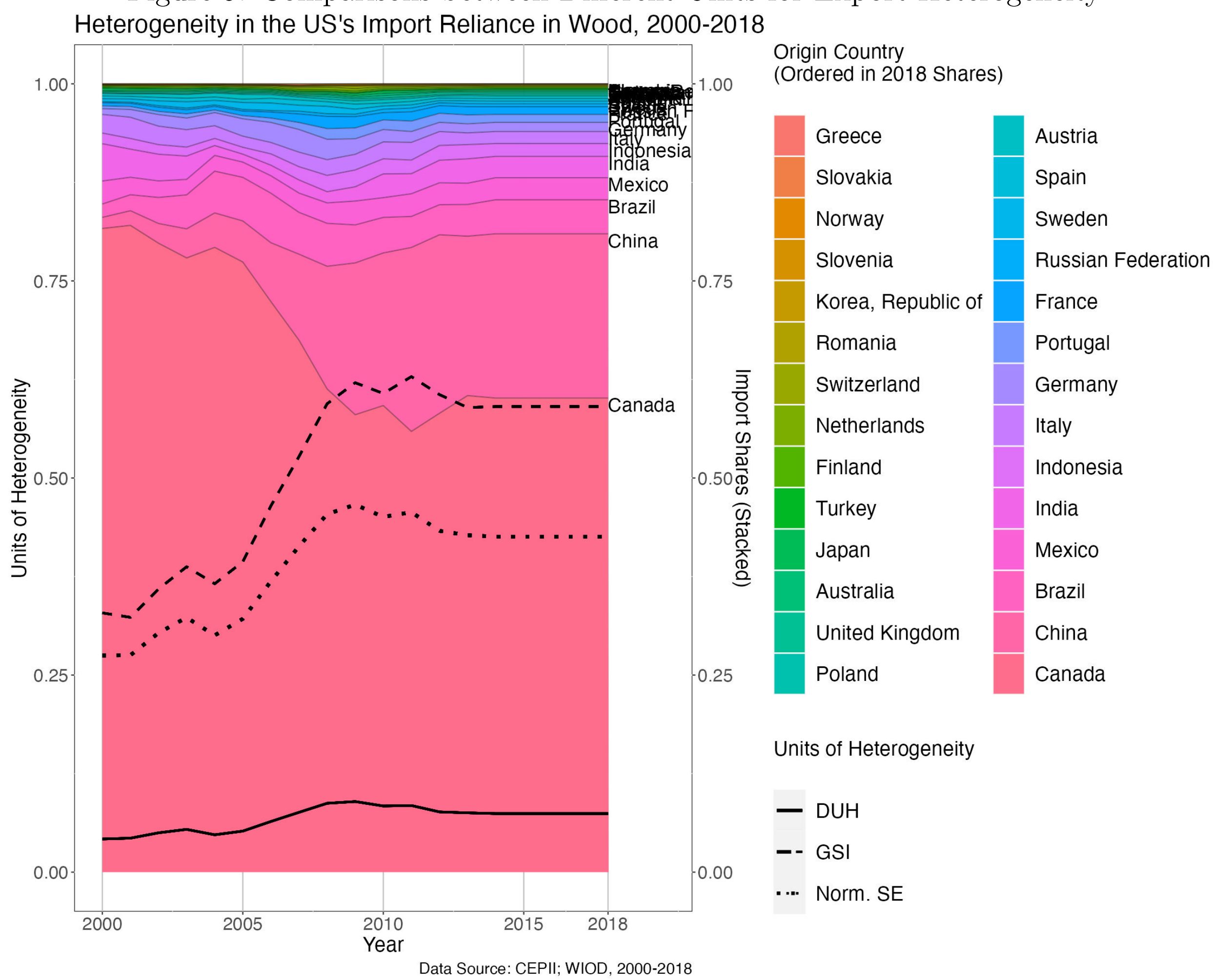
Figure 8: Comparisons between Different Units for Export Heterogeneity



[14] Destination countries are the destination of US exports. Countries that import little as well as RoW are excluded as this example aims to measure export reliance, meaning the distinct grouping of countries in the reasonable universe matters. The cut-off of top-20 is an arbitrary choice for this illustrative example.

Import Reliance.—Import reliance is not quite diametrically opposed to export reliance. If a country's import in an industry heavily relies on one country, then one can argue that said country's economic activity in that industry can be easily influenced by exogenous shocks in the origin country. Figure 9 graphs the share of the US's imports in *Wood* from 28 origin countries.[15] In figure 9 demonstrates quite well why concentration units are suitable for measuring concentration but not heterogeneity. In this example, the US's imports of wood heavily rely on Canada. The reliance is partially substituted by Chinese imports starting in 2007, but the reliance is still high. In terms of market concentration, GSI and SE reflect that the US's imports of wood became less concentrated in 2007. In terms of heterogeneity, DUH reflects that the distribution in this system is still highly homogeneous.

Figure 9: Comparisons between Different Units for Export Heterogeneity



This last example brings my discussion back to the definition of heterogeneity, and the place in the literature for a measure of it. As is evidenced here, DUH is not meant to be a

[15] Origin countries are defined as the destination of US exports. Countries belonging to RoW are excluded as this example aims to measure import reliance, meaning the distinct grouping of countries in the reasonable universe matters. The 28 countries are what is available to me in this data.

replacement/improvement of existing concentration units, it is meant to complement them when heterogeneity is the outcome of interest. Heterogeneity is about the presence of mixture in the system while concentration is about the presence of large groups. I agree with the sentiment in Chakravarty and Eichhorn (1991) that concentration units are fundamentally different from units meant to measure inequality, although I believe the difference stems from more than just EXP. The examples here show the difference between heterogeneity and concentration as well as how units made to measure them can behave differently in the same environment.

6 Conclusion

Building on the axiomatizations of the Gini coefficient, Herfindahl–Hirschman index, and Shannon's entropy, I uniquely characterize the set of units/indices/measures called the Descriptive Units of Heterogeneity. This set of units is simple to use, similar to existing units, and can fairly reflect changes in the evenness of minority groups. DUH provides a specific meaning to the sentence "S is x times more heterogeneous than S'" via the Principle of Proportional Transfers.

DUH is not meant to be a replacement for either HHI or SE. The improvement in reflecting the addition of zero-groups and the changes in minority groups may not be a desirable property in cases where truly only information from large groups should be considered. In the end, I only hope that my pursuit of this new measure is guided by a reasonable and common motivation and that future researchers, empirical or otherwise, can utilize this measure, or others similar to it, to find more insights into the evolution of heterogeneity in systems.

Bibliography

Apple and Statista (2024). Share of apple's revenue by product category from the 1st quarter of 2012 to the 1st quarter of 2024. Graph.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3):244–263.

Chakrabarti, C., Chakrabarty, I., et al. (2005). Shannon entropy: axiomatic characterization and application. *International Journal of Mathematics and Mathematical Sciences*, 2005:2847–2854.

Chakravarty, S. R. and Eichhorn, W. (1991). An axiomatic characterization of a generalized index of concentration. *Journal of Productivity Analysis*, 2:103–112.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348–361.

Gaulier, G. and Zignago, S. (2010). Baci: International trade database at the product-level. the 1994-2007 version. Working Papers 2010-23, CEPII.

Hannah, L. and Kay, J. A. (1977). *Concentration in modern industry: Theory, measurement and the UK experience*. Springer.

James, D. R. and Taeuber, K. E. (1985). Measures of segregation. *Sociological Methodology*, 15:1–32.

Kvålseth, T. O. (2022). Measurement of market (industry) concentration based on value validity. *Plos one*, 17(7):e0264613.

Laakso, M. and Taagepera, R. (1979). "effective" number of parties: a measure with application to west europe. *Comparative political studies*, 12(1):3–27.

Marshall, A. W., Olkin, I., and Arnold, B. C. (1979). *Inequalities: theory of majorization and its applications*. Springer.

Nambiar, K., Varma, P. K., and Saroch, V. (1992). An axiomatic definition of shannon's entropy. *Applied Mathematics Letters*, 5(4):45–46.

Newbery, D. (1970). A theorem on the measurement of inequality. *Journal of Economic Theory*, 2(3):264–266.

Nunes, A., Trappenberg, T., and Alda, M. (2020). The definition and measurement of heterogeneity. *Translational Psychiatry*, 10(1):299.

Pomatto, L., Strack, P., and Tamuz, O. (2023). The cost of information: The case of constant marginal costs. *American Economic Review*, 113(5):1360–1393.

Rothschild, M. and Stiglitz, J. E. (1971). Increasing risk ii: Its economic consequences. *Journal of Economic theory*, 3(1):66–84.

Rothschild, M. and Stiglitz, J. E. (1973). Some further results on the measurement of inequality. *Journal of Economic Theory*, 6(2):188–204.

Ruggles, S., Flood, S., Sobek, M., Backman, D., Chen, A., Cooper, G., Richards, S., Rodgers, R., and Schouweiler, M. (2024). IPUMS USA: Version 15.0. dataset.

Schwartz, J. and Winship, C. (1980). The welfare approach to measuring inequality. *Sociological Methodology*, 11:1–36.

Sims, C. (2010). Rational inattention and monetary economics. In Friedman, B. M. and Woodford, M., editors, *Handbook of Monetary Economics*, volume 3, pages 155–181. Elsevier, Amsterdam.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Suyari, H. (2004). Generalization of shannon-khinchin axioms to nonextensive systems and the uniqueness theorem for the nonextensive entropy. *IEEE Transactions on Information Theory*, 50(8):1783–1787.

Timmer, M., Erumban, A. A., Gouma, R., Los, B., Temurshoev, U., de Vries, G. J., Arto, I.-a., Genty, V. A. A., Neuwahl, F., Francois, J., et al. (2012). The world input-output database (wiod): contents, sources and methods. Technical report, Institue for International and Development Economics.

Appendix A Proofs

Definition 3: A function $\psi : \mathbb{R}_+^{G-1} \to \mathbb{R}_+$ is a measure of evenness in minority group distribution if it is of the following form:

$$p \in \mathbb{R}_+,\ \psi_p(S) = \psi\left(\tilde{P}_2, \ldots, \tilde{P}_G\right) = 1 - \left(\sum_{g=2}^{G} \left|\tilde{P}_g - \frac{1}{G-1}\right|^p\right)^{\frac{1}{p}}.$$

Proposition 1: Consider an index $\Phi_p = \Phi(\varphi, \psi)$ that satisfies *SYM*, *INV*, and *IND*. Holding P_1 constant, if Φ is strictly increasing in ψ_p, then $\Phi(\varphi(P_1), \psi)$ satisfies the *Principle of Diminishing Transfers* if and only if $p > 1$.

Proof of Proposition 1:

Consider two ordered systems $S = (P_1, \ldots, P_g, P_{g+1}, \ldots, P_G)$ and $S' = (P_1, \ldots, P_g - c, P_{g+1} + c, \ldots, P_G)$ where $c < \frac{P_g - P_{g+1}}{2}$. Define $\tilde{c} = \frac{c}{P_2 + \cdots + P_G}$. I want to show that $\psi(S) < \psi(S')$ and $\Phi(S) < \Phi(S')$, thus satisfying *PDT*.

Given S and S', we have

$$\psi_p(S) = 1 - \left(\left|\tilde{P}_2 - \frac{1}{G-1}\right|^p + \cdots + \left|\tilde{P}_g - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} - \frac{1}{G-1}\right|^p + \cdots + \left|\tilde{P}_G - \frac{1}{G-1}\right|^p\right)^{\frac{1}{p}}$$

$$\psi_p(S') = 1 - 1\left(\left|\tilde{P}_2 - \frac{1}{G-1}\right|^p + \cdots + \left|\tilde{P}_g - \tilde{c} - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} + \tilde{c} - \frac{1}{G-1}\right|^p + \cdots + \left|\tilde{P}_G - \frac{1}{G-1}\right|^p\right)^{\frac{1}{p}}.$$

Observe that

$$\begin{aligned}
&\psi_p(S) < \psi_p(S') \\
\iff &\left(\tilde{P}_g - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} - \frac{1}{G-1}\right|^p + C\Bigg)^{\frac{1}{p}} > \left(\tilde{P}_g - \tilde{c} - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} + \tilde{c} - \frac{1}{G-1}\right|^p + C\Bigg)^{\frac{1}{p}} \\
\iff &\left|\tilde{P}_g - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} - \frac{1}{G-1}\right|^p > \left|\tilde{P}_g - \tilde{c} - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} + \tilde{c} - \frac{1}{G-1}\right|^p.
\end{aligned}$$

Case 1: $\frac{1}{G-1} < \tilde{P}_{g+1} < \tilde{P}_g$, then

$$\begin{aligned}
&\left|\tilde{P}_g - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} - \frac{1}{G-1}\right|^p > \left|\tilde{P}_g - \tilde{c} - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} + \tilde{c} - \frac{1}{G-1}\right|^p \\
\iff &\left(\tilde{P}_g - \frac{1}{G-1}\right)^p + \left(\tilde{P}_{g+1} - \frac{1}{G-1}\right)^p > \left(\tilde{P}_g - \tilde{c} - \frac{1}{G-1}\right)^p + \left(\tilde{P}_{g+1} + \tilde{c} - \frac{1}{G-1}\right)^p \\
\iff &\frac{\left(\tilde{P}_g - \frac{1}{G-1}\right)^p + \left(\tilde{P}_{g+1} - \frac{1}{G-1}\right)^p}{2} > \frac{\left(\tilde{P}_g - \tilde{c} - \frac{1}{G-1}\right)^p + \left(\tilde{P}_{g+1} + \tilde{c} - \frac{1}{G-1}\right)^p}{2}. \\
\iff &p > 1(\text{ making the function } x^p \text{ convex}).
\end{aligned}$$

Case 2: $\tilde{P}_{g+1} < \frac{1}{G-1} < \tilde{P}_g$, then

$$\begin{aligned}
&\left|\tilde{P}_g - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} - \frac{1}{G-1}\right|^p > \left|\tilde{P}_g - c - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} + c - \frac{1}{G-1}\right|^p \\
\iff &\left(\tilde{P}_g - \frac{1}{G-1}\right)^p + \left(\frac{1}{G-1} - \tilde{P}_{g+1}\right)^p > \left(\tilde{P}_g - c - \frac{1}{G-1}\right)^p + \left(\frac{1}{G-1} - \tilde{P}_{g+1} - c\right)^p \\
\iff &\underbrace{\left(\tilde{P}_g - \frac{1}{G-1}\right)^p - \left(\tilde{P}_g - c - \frac{1}{G-1}\right)^p}_{>0} + \underbrace{\left(\frac{1}{G-1} - \tilde{P}_{g+1}\right)^p - \left(\frac{1}{G-1} - \tilde{P}_{g+1} - c\right)^p}_{>0} > 0.
\end{aligned}$$

Case 3: $\tilde{P}_{g+1} < \tilde{P}_g < \frac{1}{G-1}$, then

$$\begin{aligned}
&\left|\tilde{P}_g - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} - \frac{1}{G-1}\right|^p > \left|\tilde{P}_g - c - \frac{1}{G-1}\right|^p + \left|\tilde{P}_{g+1} + c - \frac{1}{G-1}\right|^p \\
\iff &\left(\frac{1}{G-1} - \tilde{P}_g\right)^p + \left(\frac{1}{G-1} - \tilde{P}_{g+1}\right)^p > \left(\frac{1}{G-1} - \tilde{P}_g + c\right)^p + \left(\frac{1}{G-1} - \tilde{P}_{g+1} - c\right)^p \\
\iff &\frac{\left(\frac{1}{G-1} - \tilde{P}_g\right)^p + \left(\frac{1}{G-1} - \tilde{P}_{g+1}\right)^p}{2} > \frac{\left(\frac{1}{G-1} - \tilde{P}_g + c\right)^p + \left(\frac{1}{G-1} - \tilde{P}_{g+1} - c\right)^p}{2}. \\
\iff &p > 1(\text{ making the function } x^p \text{ convex}).
\end{aligned}$$

Lastly, if $G = 3$, then case 2 is the only case, meaning the RHS of the statement can be expanded to $p \geq 1$. $\square$

Lemma 1: Any measure $\Phi(n_1, \ldots, n_G)$ of system $S = (n_1, \ldots, n_G)$ satisfies SYM if $(n_1, \ldots, n_G)$ is a vector ordered such that $n_1 \geq n_2 \geq \cdots \geq n_G$.

Proof of Lemma 1:

The proof is trivial given that the groups are ordered by size and not the label of the groups. This is a convenient consequence of defining systems as mappings from the universe of groups to a vector of numbers.

Lemma 2: If an index $\Phi(\varphi, \psi)$ of heterogeneity satisfies SYM, INV, and PPT, it is monotonically decreasing in P_1, and therefore a positive monotonic transformation of $\frac{1}{P_1}$.

Proof of Lemma 2:

Take any 2 systems of G groups $S = (n_1, n_2, \ldots, n_G)$ and $S' = (n'_1, n'_2, \ldots, n'_G)$ such that $\Phi(S) > \Phi(S')$ and that the $(n_2, \ldots, n_G) = \tilde{S} = \lambda \cdot \tilde{S}' = \lambda \cdot (n'_2, \ldots, n'_G)$, $\lambda \in \mathbb{R}_{++}$, then by *Scale Invariance*:

$$\Phi(n_1, n_2, \ldots, n_G) > \Phi(n'_1, n'_2, \ldots, n'_G) = \Phi\left(n'_1 \cdot \frac{n_S}{n'_S}, n'_2 \cdot \frac{n_S}{n'_S}, \ldots, n'_G \frac{n_S}{n'_S}\right).$$

By the *Principle of Transfers*, since $n_1 + n_2 + \cdots + n_G = n'_1 \frac{n_S}{n'_S} + n'_2 \frac{n_S}{n'_S} + \cdots + n'_G \frac{n_S}{n'_S}$,

$$\Phi(S) > \Psi(S') \iff n_1 < n'_1 \cdot \frac{n_S}{n'_S} \iff P_1 < P'_1.$$

□

Lemma 3: If an index $\Phi(\varphi, \psi)$ of heterogeneity satisfies SYM, INV, PPT, and IND, then $\Phi = \varphi \cdot \psi$.

Proof of Lemma 3:

Notice first that *Independence* trivially implies that φ and ψ must be separable. Take any system S with G groups. By the *Principle of Proportional Transfers*, it must be that $\forall \alpha \in \left[1, \frac{n_1 + \tilde{P}_2 \cdot n_1}{n - 2 + \tilde{P}_2 \cdot n_1}\right]$

$$\alpha \cdot \Phi(P_1, P_2, \ldots, P_G) = \Phi\left(P_1^\alpha, P_2 + \tilde{P}_2\left(P_1 - P_1^\alpha\right), \ldots, P_G + \tilde{P}_G\left(P_1 - P_1^\alpha\right)\right)$$

$$\iff \alpha \Phi(P_1, P_2, \ldots) = \Phi\left(P_1^\alpha, P_2', \ldots, P_G'\right) \iff \alpha = \frac{\varphi\left(P_1^\alpha\right)}{\varphi(P_1)}.$$

where $\exists \lambda \in \mathbb{R}_{++}\ s.t.\ \lambda P_g = P_g',\ \forall g \in \{2, \ldots, G\}$.

□

Proposition 2: If an index $\Phi(\varphi, \psi)$ of heterogeneity satisfies INV, IND, and PPT, then it must be $\Phi = \varphi(P_1) \cdot \psi(\tilde{P}_2, \ldots, \tilde{P}_G)$ where $\varphi(P_1) = -c \cdot log_q(P_1)$, $c \in \mathbb{R}_{++}$.

Proof of Proposition 2:

From the previous 2 lemmas, we know that $\varphi(P_1)$ must be a positive monotonic transformation of $\frac{1}{P_1}$ and that for α such that $P_1^\alpha > P_2 + \tilde{P}_2(P_1 - P_1^\alpha)$, we must have

$$\alpha = \frac{\varphi(P_1^\alpha)}{\varphi(P_1)}.$$

Notice that the only positive monotonic transformation that would satisfy this is $log_q\left(\frac{1}{P_1}\right)$, up to a positive scalar multiplication. Further notice that any $log_q\left(\frac{1}{P_1}\right)$ can be rewritten as $\frac{ln\left(\frac{1}{P_1}\right)}{ln(q)}$, so it is equivalent to write $c \cdot ln\left(\frac{1}{P_1}\right)$. As such, $\varphi(P_1) = c \cdot ln\left(\frac{1}{P_1}\right)$, $c \in \mathbb{R}_{++}$ is the unique function, up to positive scalar multiplication, of majority proportions that can lead to $\Phi(\varphi, \psi)$ satisfying, *Independence*, *Scale Invariance*, and *Principle of Proportional Transfers*. □

Theorem 1: The *Descriptive Units of Heterogeneity* Φ defined as:

$$\Phi_p(n_1, \ldots, n_G) = -\frac{ln\left(P_1\right)}{ln(G)}\left[1 - \left(\sum_{g=2}^{G}\left(\tilde{P}_g - \frac{1}{G-1}\right)^p\right)^{\frac{1}{p}}\right]$$

$$= \frac{ln\left(P_1\right)}{ln(G)}\left[\left(\sum_{g=2}^{G}\left(\tilde{P}_g - \frac{1}{G-1}\right)^p\right)^{\frac{1}{p}} - 1\right].$$

where $P_1 = \dfrac{n_1}{n_1 + \cdots + n_G}$, $\tilde{P}_g = \dfrac{n_g}{n_2 + \cdots + n_g}$, $p \in (1, \infty)$.

is the unique class of units that satisfy *Scale Invariance*, *Group Symmetry*, *Independence*, *Principle of Diminishing Transfers*, *Principle of Proportional Transfers*, *Contractibility*, and uses ψ_p to account for evenness in minority.

Proof of Theorem 1:

Propositions 1 and 2 combined implies that DUH satisfies *SYM*, *INV*, *PPT*, *IND*, and *CON*, but not necessarily *PDT*. I have only shown that DUH satisfies *PDT* when either φ or ψ_p is held constant, meaning I need to be careful with the functional form of $\varphi(P_1)$. To satisfy *PDT* overall, it must be that when there is a transfer from the majority group to the minority group, the increase in Φ through $\varphi(P_1)$ dominates the decrease in Φ through ψ_p in the case where evenness decreases as a result of the transfer.

Notice that to show Φ satisfies *PDT*, we only need to look at the extreme case where P_1 is close to 1 and $\psi = 1$. In this case, a simple transfer from n_1 to n_2 will decrease ψ_p the most. For simplicity, we will consider the case when $p = 2$ so that Φ is simply:

$$\Phi(n_1, \ldots, n_G) = \frac{ln\,(P_1)}{ln(G)} \left[\sqrt{\sum_{g=2}^{G} \left(\tilde{P}_g - \frac{1}{G-1} \right)^2} - 1 \right].$$

Denote $n_2 + \cdots + n_G$ as $\tilde{n}_S$, a transfer of x from n_1 to n_2 when $\psi = 1$ can be written as:

$$\Phi_2 = ln\left(\frac{n_1 - x}{n_S}\right) \left[\sqrt{\left(\frac{\frac{\tilde{n}_S}{G-1} + x}{\tilde{n}_S + x}\right) + (G-2)\left(\frac{\frac{\tilde{n}_S}{G-1}}{\tilde{n}_S + x} - \frac{1}{G-1}\right)^2} - 1 \right].$$

Taking the derivative of this expression with respect to x, we have, $\forall x \in \left[0, \frac{(G-1)n_1 - \tilde{n}_S}{G}\right]$:

$$\frac{d}{dx}\Phi_2(n_1 - x, n_2 + x, n_3, \ldots, n_G) = \frac{\sqrt{\frac{G-2}{G-1}}\left[\tilde{n}_S(x - n_1)ln\left(\frac{n_1 - x}{n_S}\right) + x(b + x)\right]}{(x - n_1)(x + \tilde{n}_S)^2} + \frac{1}{n_1 - x} > 0.$$

$\square$

Appendix B Characterization Axioms of HHI and SE

[EXP] Expandability. $\Phi(n_1, \ldots, n_G)$ satisfies *Expandability* if

$$\Phi(n_1, \ldots, n_G, 0) = \Phi(n_1, \ldots, n_G).$$

It should be clear that neither EXP nor CON attempts to pin down the functional form of a unit. Rather, these two opposing axioms serve as the divide between a unit for concentration and a unit for heterogeneity.

As discussed extensively in Atkinson (1970), the fact that PDT only induces partial ordering implies that specific functional forms can always be chosen to induce different total orders when neither system's distribution second-order stochastically dominates the other.[16] The functional forms of both HHI and SE were able to be uniquely characterized because of this feature.[17] HHI uses EXP and the *Replication Principle* (REP) and SE uses EXP and *Shannon's Additivity* (SADD).

[REP] Replication Principle. $\Phi(n_1, \ldots, n_G)$ satisfies the *Replication Principle* for concentration if replicating a system k times divides the system concentration by k.

For example, take $k \in \mathbb{N}$,

$$\frac{1}{k}\Phi(n_1, \ldots, n_G) = \Phi\left(\underbrace{\frac{n_1}{k}, \frac{n_1}{k}, \ldots, \frac{n_1}{k}}_{\text{Sum to } n_1}, \frac{n_2}{k}, \frac{n_2}{k}, \ldots, \underbrace{\frac{n_G}{k}, \ldots, \frac{n_G}{k}}_{\text{Sum to } n_G}\right).$$

REP pins down the cardinal meaning of the unit by linking the multiplication of the unit to how many times a system is divided/replicated into a system with more groups. Chakravarty and Eichhorn (1991) shows that if a concentration unit C can be represented as a *self-weighted quasilinear mean*, then C is the Hannah-Kay index of concentration if and only if C satisfies the fundamental axioms and REP.[18] Since HHI is $H^n_{\alpha=2}(S)$, its is the unique self-weighted quasilinear concentration unit that satisfies the fundamental axioms, EXP, and REP. Notice REP applies only when *every group* is broken into multiple groups. To prescribe how the index behaves when only one group is broken up, SE forgoes REP and restricts on *Shannon's Additivity* instead.

[16] Newbery (1970) shows that no additive functional forms can be chosen to induce the same order as the Gini coefficient. Such impossibility theorem stems from and reaffirms the point made by Atkinson (1970).

[17] Similarly, I utilize this feature to uniquely characterize the Descriptive Units of Heterogeneity.

[18] A relative concentration index $C : D \to \mathbb{R}$ is called a self-weighted quasilinear mean if for all $n \in \mathbb{N}$, $x \in D^n$, $C^n(x)$ is of the form:

$$C^n(x) = \phi^{-1}\left[\sum_{g=1}^{G} P_g \phi(P_g)\right],$$

where $\phi : (0, 1] \to \mathbb{R}$ is strictly monotonic.

[SADD] Shannon's Additivity. Define $n_{gj} \geq 0$ such that $n_g = \sum_{j=1}^{m_g} n_{gj}$, $\forall g \in \{1, \ldots, G\}$, $\forall j \in \{1, \ldots, m_g\}$

$\Phi(n_1, \ldots, n_G)$ satisfies *Shannon's Additivity* if

$$\Phi(n_{11}, \ldots, n_{Gm_G}) = \Phi(n_1, \ldots, n_G) + \sum_{g=1}^{G} \frac{n_g}{n_S} \cdot \Phi\left(\frac{n_{g1}}{n_g}, \ldots, \frac{n_{gm_g}}{n_g}\right).$$

which implies (by setting $m_g = 1$, $\forall g \in \{1, \ldots, G-1\}$ and $n_{G'} = n_G + n_{G+1}$),

$$\Phi(n_1, \ldots, n_G, n_{G+1}) = \Phi(n_1, \ldots, n_{G'}) + \frac{n_{G'}}{n_1 + \ldots + n_{G-1} + n_{G'}} \cdot \Phi\left(\frac{n_G}{n_{G'}}, \frac{n_{G+1}}{n_{G'}}\right).$$

SADD pins down how the decomposition of group(s) in a system should influence the unit.[19]

For detailed proofs of the unique characterization of SE as well as explanations of SADD, readers should refer to Suyari (2004) and Chakrabarti et al. (2005).

[19] An example of decomposing a group is to split the sales of Mac into Mac desktops and Mac laptops, for the purpose of measuring the heterogeneity of Apple's revenue streams.