Descriptive Units of Heterogeneity: An Axiomatic Approach to Measuring Heterogeneity

Willy Chen*

2024

Abstract

This paper addresses the challenge of measuring between-group heterogeneity in systems, as existing measures lack cardinal interpretation and comparability across systems/population. Using an axiomatic approach, I highlight the strengths and limitations of existing measures and generalize properties that need to be satisfied by better alternatives. Using these axioms, I propose the class of measures called the *Descriptive Units of Heterogeneity* (DUH), a hybrid solution to prior limitations without limiting the applicable contexts. DUH achieves the generalized comparability of concentration units while still able to reflect changes in the distribution of small groups in the population. Hence, DUH provides a valuable tool for empirical researchers studying heterogeneity in systems in various contexts, such as racial composition in a city or revenue shares by products of a firm.

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

*Michigan State University: willyc@msu.edu

1 Introduction

Measuring heterogeneity is, at its heart, a dimension-reduction problem. When studying a complex population, one hopes to simplify the complexity without losing the big picture. Deciding the essential elements for *describing* a population is the key to defining a measure that can be useful for empirical researchers.

My objective is to measure the heterogeneity present in a system – a collection of distinct groups with more than one elements. For example, consider the revenue streams by each Apple product a group, the sales number for each product is population/number of elements, in that group. The collection of these groups—iPhone, iPad, Mac, Services, and Wearables & Accessories—is the system, Apple's revenue streams.

The heterogeneity of a system is commonly measured in two ways: Dispersion units and Concentration units. Dispersion units focus on measuring the distance between the observed population distribution to a benchmark distribution, yielding concise interpretations at the cost of comparability between systems. Concentration units focus on measuring the richness of information from select subgroups of the population. These units differ from dispersion units by having generalized compatibility between systems, but they understate the information provided by small subgroups.

In this paper, I propose a new yet intuitive way to think about the make-up of heterogeneity by separating it into *contribution by the largest group* (majority) and *evenness of the rest of the groups* (minority) and show that this, in fact, builds on the existing paradigm. I start by discussing the *equivalence axioms*, axioms that pin down when two systems of the same number of groups are equally heterogeneous, and the *monotonicity axioms*, axioms that pin down how two systems of the same number of groups but are marginally different should be ordered. I then show that *Herfindahl–Hirschman Index* (HHI) and *Shannon's Entropy* (SE) satisfy the existing axioms, and are further uniquely characterized by their own sets of axioms. I define new axioms where the existing ones fall short and detail how these axioms keep the generalized comparability between systems while still accounting for information provided by small subgroups. These axioms together characterize a class of indices that focus on making the comparisons between system *descriptive* and this class of measures are termed the *Descriptive Unit of Heterogeneity.*

I use changes in racial composition in San Francisco and changes in Apple's revenue shares as empirical examples to illustrate the strength of my index in accounting for both the contribution of the majority groups and the evenness of the minority groups, a feature and a curse of the concentration indices by design.

The rest of the paper proceeds as follows. Section 2 provides motivation for a new paradigm using features of select existing units of heterogeneity. Section 3 lists the *Equivalence Axioms* and *Monotonicity Axioms*, followed by my proposed paradigm of thinking about heterogeneity, and then the *Decomposition Axioms* which uniquely characterizes different units of heterogeneity. Section 4 defines the *Descriptive Unit of Heterogeneity* and compares this new index to the existing concentration units. Section 5 presents empirical examples to highlight the strength of the descriptive unit of heterogeneity in various settings.

This paper concludes with a summary of the differences between my index and existing ones and I generalizes best practices with using this index. I use the appendix to present proofs necessary for establishing uniqueness of my index.

2 Lessons from Existing Units of Heterogeneity

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

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, let S be Michigan State University, 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, i.e., 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.

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

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

Definition 2: A system S' of G groups is said to achieve **minimum heterogeneity/perfect homogeneity** if it can be represented as a vector of size G where all but one entry are 0:

$$S' = (0, 0, \dots, 0, n, 0, \dots, 0) = n \cdot (0, 0, \dots, 0, 1, 0, \dots, 0)$$

The one-dimensional (presence of mixture) nature of this definition makes it convenient for any measure to be bounded between $\Phi(S) \in \mathbb{R}_{++}$ and $\Phi(S) = 0$. The harmlessness of this generalization is evidenced by existing measures of heterogeneity, even when they are constructed with different goals in mind. These units can be generally separated into two categories - Dispersion units and Concentration units.

This section provides brief discussions of select dispersion and concentration units that motivates the formalization of axioms in section 3. I will show that even though dispersion units often have limited use cases, the direct inference they enable make them an attractive option. On the other hand, concentration units excels at generalizing comparisons between systems while they focus on indirect inferences limited to the majority in a system. The axioms discussed in section 3 thus are the products of my desire to find an index that can have the strength of both types of units while suffering minimally from potential limitations.

2.1 Dispersion Units (Gini Coefficient and Dissimilarity Index)

If you ever venture into the literature of economics of inequality, you likely would have heard of the Gini coefficient or the Dissimilarity Index, both of which are dispersion units.

Gini coefficient standardizes the population to a single and ordered outcome such as household income. It compares the distribution of said outcome to the uniform distribution. The idea is simple, but powerful. Because the outcome is standardized, it is not hard to understand the sentence "the top x% of households in the US earns the top y% of income." In a uniformly distributed world, x and y should be equal, and if it is not, then there is inequality.

Part (a) in Figure 1 is how this is taught in most undergraduate development classes. The Gini coefficient of a country is defined as $\frac{A}{A+B}$, and the higher the area of A, the more inequality there is. Something those future development economists may not understand, is that (1) the line separating areas A and B is called the Lorenz curve, and (2) using Gini coefficients of two countries to make inference about inequality implicitly assumes the *Lorenz Criteria* - that the two Lorenz curves do not cross (Schwartz and Winship 1980; James and Taeuber 1985).

For example, if we look at part (b) of Figure 1, we can compare a country with $B + C + D$ under the Lorenz curve to a country with just B under the Lorenz curve, and say that the second country has more inequality (because A is smaller than $A + C + D$). But we cannot compare a country with $B + C$ under the Lorenz curve to a country with $B + D$ under the Lorenz curve, and say either country has more inequality than the other.[1]

Dissimilarity index does not have quite the restriction that Gini coefficient does, or at least we can be clear about the limitations before drawing its version of the Lorenz curve. The idea of dissimilarity index is simple, yet elegant - If I know the overall distribution of **two groups** in a city, I can know if there is sorting, i.e., dissimilarity between the neighborhoods. By summing up the differences between each neighborhood's observed distribution and its non-sorted distribution (same as the overall city), we can infer higher dissimilarity from just a higher sum. For example, we can write dissimilarity index for income sorting in a city with I neighborhoods as:

$$DI = \frac{1}{2}\sum_{i=1}^{I}\left|\frac{\#LowIncome_i}{\sum_{i=1}^{I}\#LowIncome_i} - \frac{\#NotLowIncome_i}{\sum_{i=1}^{I}\#NotLowIncome_i}\right|$$

[1] Schwartz and Winship (1980) provides several excellent examples an explanation of why the Lorenz criteria significantly limits use cases of the Gini coefficient. I omit this discussion as that is not the focus of this paper.

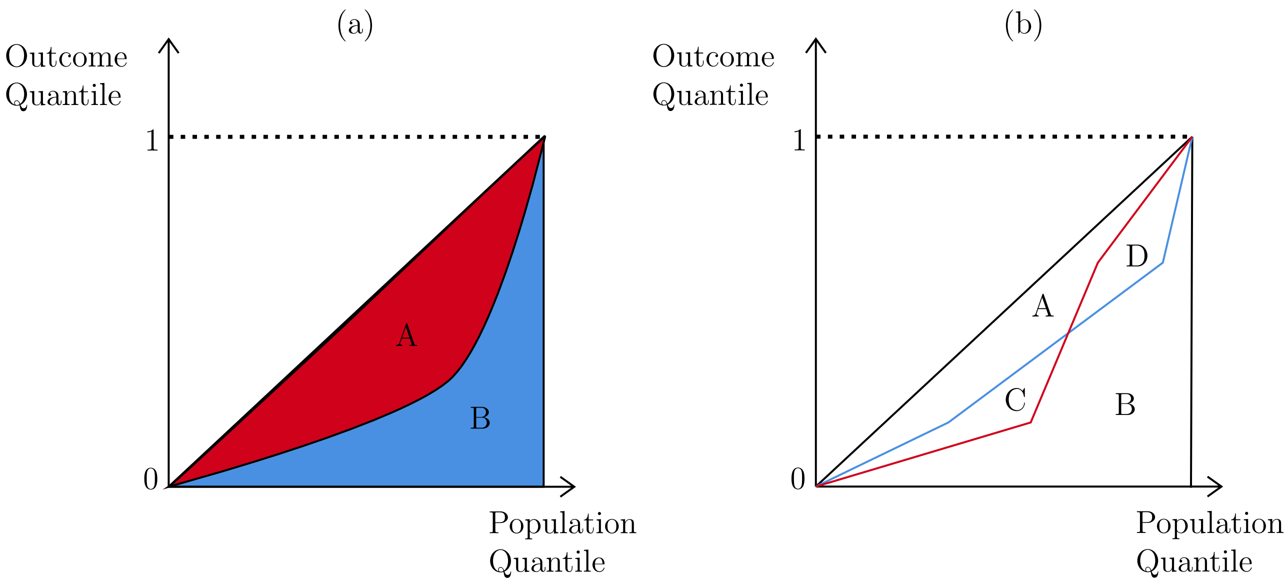


Figure 1: Gini Coefficient: Definition and Comparison

This index has been useful in the broad literature of income/racial sorting, but it does suffer from one design flaw - there can only be two groups. It is straightforward to interpret an increase in DI when it is made up of differences in proportions of a binary outcome, but the nature of differences between proportions obscures natural extensions to systems of three or more groups.

These are not the only dispersion units, but they are excellent examples of indices that are useful because of their straightforward interpretation, despite the limitations discussed. Overall, comparing the observed distribution to a benchmark distribution is intuitive and easy to understand. But the ways these are designed make them less practical in modern empirical research where we may need general comparison between systems. Concentration units, on the other hand, suffers from just the opposite.

2.2 Concentration Units (Herfindahl–Hirschman Index and Shannon's Entropy)

A simple escape from the above mentioned limitations is to use the Concentration units. Let there be a system S with G groups. The *Herfindahl–Hirschman Index* (HHI) and its complement Gini-Simpson Index of system S are defined as:

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

HHI, and subsequently GSI, has an alluringly simple interpretation - The probability of 2 random draws with replacement, from the system S, being from the same group in S (in this case, GSI is quite literally the complement of that). The elegance of this measure had historically outshined its flaws, perhaps because it is developed for contexts where such flaws were intentional.

HHI disproportionately accounts for changes in large groups. Observe that, due to the squaring of group proportions, any group accounting for less than, say 10%, of the system has fewer than 1% impact on the system's HHI. For example, say that MSU in 2020 has 80% students, 15% staff, and 5% faculty, and that those respective numbers change to 80%, 19%, and 1% in 2030. The respective HHIs are 0.665 and 0.6762. In this case, the faculty population shrank by 80%, but the changes in HHI is barely noticeable.

This was not previously an issue as HHI is most used in measuring market share/power, so a 80% shrinkage of a firm that was only at 5% before may not mean much if those 4% went to another small firm. However, this means that HHI should be used only in context that are similar to discussing market shares and market concentrations, opposite of a unit for measuring heterogeneity.

HHI's simplistic interpretation makes it hard to back out of the original distribution. Unlike the dispersion units discussed above, when a system's HHI increases by some number x, the only interpretation is the probabilistic one, which is seldom descriptive in the case of measuring heterogeneity. Combining with the fact that it accounts mostly for the larger groups, using HHI costs researchers much information about movements within the smaller groups. Two systems, $(48\%, 48\%, 4\%)$ and $(60\%, 30\%, 10\%)$ will yield 0.4624 and 0.46 HHI. When we look at the proportions, we can see that the difference is big, but that is hardly reflected in the difference in HHI.

Also suffering from this flaw is the commonly used measure of uncertainty in information theory - Shannon's Entropy, defined as:

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

One key thing to notice is that in both of these units, zero-groups, groups with zero elements ($n_g = 0$), 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. One likely does not care about whether Starbucks has any market power in the pencil lead industry, and one need not care about whether the probability of finding pencil lead in their coffee is accounted for or not. But why should the same be said about any measure of heterogeneity?

If I want to draw a valid comparison of heterogeneity between two systems, they should be similar enough for me to compare. And if I were to compare them, I should only compare them along the dimensions that make sense for the context. Comparing apples to oranges might not necessarily be nonsensical like the popular idiom suggests. In the proper context, I can compare an apple to an orange on density, brightness of color, amount of sugar per ml of water, etc. I can even compare an apple to an orange on how good each fruit is at being a citrus (clearly, the apple will lose). What I cannot do is say that an apple is x times denser than an ice cube than an orange is y times rounder than a bowling ball.

To effectively compare heterogeneity between two systems, it is crucial to establish the qualifiers that make them comparable. When assessing system complexity, considering all groups, including 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. "*Zero output independence means that the index value does not alter if there is addition or deletion of a firm with zero output. This particular principle shows the fundamental difference between indices of concentration and indices of inequality. It is generally agreed that adding (deleting) an individual with zero income to (from) a population increases (decreases) inequality. But firms which produce no output should not have any impact on concentration*" (Chakravarty and Eichhorn 1991).

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

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

This index improves system compatibility by accounting for zero-groups via the normalization, but it is done at the cost of HHI's simple probabilistic interpretation. Consider the following two systems S and S':

$$S = (0.4, 0.4, 0.2)$$
$$S' = (0.5, 0.3, 0.1, 0.1)$$

These two systems have the same HHI (0.36), but they have vastly different NHHIs ($NHHI(S, G = 3) = 0.04$ and $NHHI(S', G = 4) \approx 0.15$). 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.

2.3 The Shopping List for An Unknown Recipe

Learning from the strengths and weaknesses of these units, I want to create an index that can yield simple interpretations when comparing any two systems without giving up too much for complexity. The following list are desired features that either directly or indirectly motivate the axioms that my index Φ should satisfy:

- A $\Phi(S') = x \cdot \Phi(S)$ can be interpreted as S' is x times as heterogeneous as S via some channel.
- Φ and existing units/indices share some basic properties.

- Φ accounts for the presence of zero-groups.

With these in mind, I will discuss the axiomatization of measures of heterogeneity.

3 Axioms for Indices of Heterogeneity

This paper is not a first attempt in this strand of literature at an axiomatic characterization measure. In fact, there is abundant existing literature that axiomatize Gini coefficient (Schwartz and Winship 1980; James and Taeuber 1985), HHI (Kvålseth 2022; Chakravarty and Eichhorn 1991), and SE (Nambiar et al. 1992; Suyari 2004; Chakrabarti et al. 2005). Unsurprisingly, axioms in these works have much overlap. They can be generally separated into three categories:

1. *Equivalence Axioms:* Axioms that pin down when two systems are equally heterogeneous
2. *Monotonicity Axioms:* Axioms that pin down when two systems can be ordered in their heterogeneity
3. *Decomposition Axioms:* Axioms that pin down the effects of adding/removing a group from a system on the index

3.1 Equivalence Axioms

[SYM] Group Symmetry. Given the same number of groups (G) in the system, the heterogeneity of a system is invariant to relabelling groups.

$\forall n_a, n_b, n_c \in \mathbb{N}$,

G. \ S.	S1		S2		S3
A	n_a		n_b		n_c
B	n_b	$\sim$	n_a	$\sim$	n_b
C	n_c		n_c		n_a

SYM is an intuitive axiom as it enables the index to focus on the distribution over groups in a system, rather than 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 universe have the same number of groups.

In a similar generalization effort, any index of heterogeneity should focus on the proportion (relative sizes) of each group and not the absolute sizes of each group. This generalization yields this next axiom - *Scale Invariance.*

[INV] Scale Invariance. Given two systems with the same proportions in each group, the heterogeneity of the combined system must be the same. Without loss of generality, this can be generalized to the following:

$\forall n_a, n_b, n_c, \in \mathbb{N},\ \lambda \in \mathbb{R}_{++},$

G. \ S.	S1		S2		S1+S2
A	n_a		λn_a		$(1+\lambda)n_a$
B	n_b	$\sim$	λn_b	$\sim$	$(1+\lambda)n_b$
C	n_c		λn_c		$(1+\lambda)n_c$

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

The equivalence axioms give us a convenient way to equate heterogeneity between systems with different sizes but the same distribution. Having generally pin down heterogeneity equivalence between systems, my next task is to pin down the order between systems with unequal heterogeneity. This necessitates the set of axioms called *monotonicity axioms.*

3.2 Monotonicity Axioms

Monotonicity axioms are the set of axioms that pin down the ordering of systems with the same number of groups when they are not equal.

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

Take any $n_a, n_b, n_c, n_d \in \mathbb{N}$ such that $n_a > n_b > n_c > n_d$ and let $\varepsilon < \min\left\{n_c - n_d,\ n_b - n_c,\ \dfrac{n_a - n_b}{2}\right\}$, then

G. \ S.	S1		S2		S3		S4
A	n_a		$n_a - \varepsilon$		$n_a - \varepsilon$		$n_a - \varepsilon$
B	n_b	$\prec$	$n_b + \varepsilon$	$\prec$	n_b	$\prec$	n_b
C	n_c		n_c		$n_c + \varepsilon$		n_c
D	n_d		n_d		n_d		$n_d + \varepsilon$

When comparing distributions using quantiles, the most common use case, assuming *SYM*, *INV*, and *PDT* is equivalent to assuming the *Lorenz Criterion* - the Lorenz curves of two distributions do not cross. However, the general case of the *Lorenz Criterion* is only equivalent to *PDT*. Equivalently, intuitive as *PDT* may seem, it only yields partial ordering of distributions (Rothschild and Stiglitz 1969; Atkinson et al. 1970; Rothschild and Stiglitz 1973; Kolm 1976). The implementation of *PDT* in a measure of inequality is well discussed in mathematics and statistics as *majorization*. Chapter 1 of Marshall et al. (1979) has an excellent discussion on the limits of using majorization to study inequality, and how further assumptions/axioms are necessary to propel the indices of inequality in practical use cases.

A New Way to Think About Heterogeneity

Before I continue in the world of axiomatization, I want to propose a new way to think about heterogeneity. All the units I have discussed up to this point treat each group in the system numerically equally, partially in order to ensure *SYM*. However, group symmetry can be satisfied without each group getting identical numerical treatment. Rather, it can be satisfied simply by giving the same *types* of groups the same numerical treatment.

Notice that for a index that can be used for generalized comparison between systems, it needs to be able to order the heterogeneity of systems of any positive integer number of groups. Notice that for any system, there is always the largest group and the remaining groups. What I propose is an index that treats this largest group differently than the rest of the groups in the system. The advantage to this new way of thinking about heterogeneity is that if the influence of the largest group on the index can be orthogonal to the influence of the rest of the groups, we can equate any changes in the one-dimensional index with the equivalent change(s) in either group(s) while holding the other constant.

By *SYM*, groups in each system can be ordered by the size of each group, meaning that giving the largest group a different treatment than the rest of the groups does not violate group symmetry. For convenience, I will now call the largest group the *majority group* and the rest of the groups the *minority groups*.

Under this new paradigm, I can refine the definition of a unit of heterogeneity. Recall our earlier definition - The heterogeneity of system S is measured by $\Phi : S \to \mathbb{R}_+$ such that

$$\Phi(n_1, \ldots, n_G) \geq \Phi(n'_1, \ldots, n'_G) \iff S \text{ is weakly more heterogeneous than } S'$$

Now consider the paradigm where we can separate what matters in a system into two parts:

1. Relative size of the Majority Group: $P_1 = \frac{n_1}{n_1 + \cdots + n_G}$
2. Relative size(s) of the Minority Group(s): $P_2, \ldots, P_G$

An index for heterogeneity is thus $\Phi = \Phi(\varphi, \psi)$ where φ is the influence of the relative sizes of the majority groups and ψ the influence of the relative sizes of the minority groups. If we can find well-defined functions ϕ and ψ, we can then order any two systems that we deem comparable in the context of our work. The rest of the axioms will focus on refinements of *PDT* for gaining complete ordering of system heterogeneity in an intuitive way.

For the influence of P_1 and P_2 through P_G to be orthogonal, consider the *Independence Axiom.*

[IND] Independence. The influence of the relative size of the majority population on the unit of heterogeneity should be independent of the relative sizes of the minority groups, and vice versa.

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

To satisfy *IND*, we need a function that omits P_1 by design and does not violate any of the previous axioms. Let $\tilde{P}_g = \frac{n_g}{n_2+\cdots+n_G}$, $\forall g \in \{2, \ldots, G\}$,

$$\tilde{P}_g = \frac{n_g}{n_2 + \cdots + n_G} = \frac{\frac{n_g}{n_1+\cdots+n_G}}{\frac{n_2}{n_1+\cdots+n_G} + \cdots + \frac{n_G}{n_1+\cdots+n_G}} = \frac{P_g}{P_2 + \cdots + P_G}$$

Following the similar methodology in dispersion units for easy interpretations, I want ψ to reflect the evenness in the minority groups distribution. Specifically, I define ψ to be the distance between the observed distribution in the minority groups and the ideal uniform distribution in the minority groups. This means that ψ can be characterized as a class of functions by all finite p-metrics d_p in $\mathbb{R}^n$.

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: $p \in (1, \infty)$,

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

Proposition 1: Consider an index $\Phi = \Phi(\varphi, \psi)$ that satisfies *SYM*, *INV*, and *IND*. Holding P_1 constant[2], if ψ is as defined above and Φ is strictly increasing in ψ, then $\Phi(\varphi(P_1), \psi)$ satisfies the *Principle of Diminishing Transfers* if and only if $p > 1$.

Sketch of Proof: Given the absolute value function, the proof simply follows the mechanics of Jensen's inequality. The full proof is included in the appendix.
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.*

[2]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 majority group to the minority group, the increase in $\varphi(P_1)$ must dominate the decrease in ψ_p in the case where evenness decreases.

[PPT] Principle of Proportional Transfers.

Holding the order of groups constant, a transfer from the majority group proportionally to the minority groups that reduces P_1 to P_1^α increases heterogeneity by a factor of α.

G. \ S.	S1		S2
θ_1	P_1		P_1^α
θ_2	P_2	$\sim$	$P_2 + \frac{P_2}{P_2+P_3}(P_1 - P_1^\alpha)$
θ_3	P_3		$P_3 + \frac{P_3}{P_2+P_3}(P_1 - P_1^\alpha)$

Then

$$\Phi(S_1) < \alpha\Phi(S1) = \Phi(S2)$$

PPT gives changes in the index a simple interpretation. If $2\Phi(S1) = \Phi(S2)$, then one can say that $S2$ *is twice as diverse as* $S1$ because it has the equivalent heterogeneity as if the majority group proportion in $S1$ shrunk by the power of 2 while still being the majority group, holding the same evenness in the minority groups.

3.3 Decomposition Axioms

Decomposition axioms pin down the changes to an unit when a group (zero or otherwise) is added to the system. In existing units of heterogeneity, decomposition axioms often uniquely characterize certain units to fit in the environment and to give the units a sense of cardinal meaning. The first axiom of note is the axiom of *Expandibility*, an axiom that is favored by concentration units.

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

$$\Phi(n_1, \dots, n_G) = \Phi(n_1, \dots, n_G, 0)$$

As discussed in Chakravarty and Eichhorn (1991), *EXP* is a salient reason why concentration units such as HHI and SE should not be used to measure inequality exactly because the information provided by the presence of zero-groups is not pertinent when measuring concentration, but it would be unreasonable to say the same for a unit meant to measure inequality in and heterogeneity of a system.

As a contrary axiom to *EXP*, I propose the axiom of *Contractibility*.

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

$$\Phi(n_2, \dots, n_G, 0) < \Phi(n_2, \dots, n_G)$$

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 et al. (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 dominates the other. Both HHI and SE[3], use this feature to motivate axioms for functional forms to uniquely characterize the measure. HHI uses *EXP* and the *Replication Principle* (REP) and SE uses *EXP* and the *Shannon's Additivity* (SADD).

Instead of focusing on adding or subtracting a single group from the system, *REP* pin down the cardinal meaning of the unit by linking the multiplication of the unit to the how many times a system is divided. On the other hand, *SADD* pin down how exact decompositions of group(s) in a system should influence the unit. As shown in Chakravarty and Eichhorn (1991) and Chakrabarti et al. (2005), the exponentiated version of SE satisfies *REP* but not *SADD*. These two axioms are detailed below, and I provide a simple sketch of how the units are uniquely characterized.

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

$\forall k \in \mathbb{N}$,

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

Chakravarty and Eichhorn (1991) provide the detailed steps to uniquely characterize HHI as a special case of the *Hannah-Kay* class of concentration index, with perception α and $n \in \mathbb{N}$ firms in industry $x \in D^n$, defined as:

$$H^n_\alpha(x) = \begin{cases} \left[\sum_{i=1}^{n} s_i^\alpha\right]^{\frac{1}{\alpha-1}} & \text{if } \alpha > 0,\ \alpha \neq 1 \\ \prod_{i=1}^{n} s_i^{s_i} & \text{if } \alpha = 1 \end{cases}.$$

Chakravarty and Eichhorn (1991) shows that a concentration index C can be represented as a *self-weighted quasilinear mean*[4]. Then C is the Hannah-Kay index of concentration if and only if C satisfies the replication principle.

[3] and later on, the *Descriptive Units of Heterogeneity*

[4] A relative concentration index $C : D \to \mathbb{R}$ is called a **self-weighted quasilinear mean** if for all $n \in$

[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 $n_{G'} = n_G + n_{G+1}$),

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

For detailed steps as well as explanations of *SADD*, readers should refer to Suyari (2004) and Chakrabarti et al. (2005).

Remark: The HHI discussed in this paper is the case of the HK concentration index with $\alpha = 2$. The SE discussed in this paper is the negative of the natural log of the HK concentration index with $\alpha = 1$.

4 The Descriptive Units of Heterogeneity

Using the lessons learned from the other units of heterogeneity, I propose the *Descriptive Units of Heterogneeity*—a class of units that balances interpretability and comparability. Let $n_1 \geq n_2 > 0$, $P_1 = \frac{n_1}{n_1+n_2+\cdots+n_G}$, and $\tilde{P}_g = \frac{n_g}{n_2+\cdots+n_G}$.

The Descriptive Units of Heterogeneity (DUH) of system S with $G \geq 2$ groups is defined as:

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

Theorem 2[5]: The *Descriptive Units of Heterogeneity* Φ 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*, and *Contractibility*.

This table outlines the axioms discussed and whether Gini, DUH, HHI, or SE satisfy them:

$\mathbb{N}$, $x \in D^n$, $C^n(x)$ is of the form:

$$C^n(x) = \phi^{-1}\left[\sum_{i=1}^{n} s_i \phi(s_i)\right].$$

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

[5]Proof of Theorem 2 is in the appendix.

Type	Axiom	Gini	DUH	HHI	SE
Equivalence	Type Symmetry	✓	✓	✓	✓
	Scale Invariance	✓	✓	✓	✓
Monotonicity	Principle of Diminishing Transfers	✓	✓	✓	✓
	Independence	×	✓	✓	✓
	Principle of Proportional Transfers	×	✓	×	×
Decomposition	Expandability	×	×	✓	✓
	Contractibility	✓	✓	×	×
	Replication Principle	×	×	✓	×
	Shannon's Additivity	×	×	×	✓

DUH is similar to what Gini coefficient should look like with generalized comparability across systems made up of discrete and unordered groups. DUH builds on the partial ordering of Gini coefficient to induce a total order that enables better comparisons across systems that do not satisfy the Lorenz criterion by further refining *PDT*. DUH is characterized differently than concentration units because it focuses on the overall distribution without losing comparability. In the next section, I provide empirical examples that highlight the strength of DUH.

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

The implication of this paradigm is that the set Θ needs to be handled with care, because each $\theta \in \Theta$ must be similar/comparable to each other. Figures 2 and 3 illustrates this idea with a practical example. These two figures present racial composition of San Francisco MSA from 2007 to 2022 using ACS 1-year data (Ruggles et al. 2024). Figure 2 defines Θ as $\{$White, Black, Other$\}$ while figure 3 splits up the *Other* group into 3 sub-groups, yielding $\Theta = \{$White, Black, Asian, Native America, Multi-Race$\}$.

In figure 2, the heterogeneity of this system is somewhat stable due to the influence of the shrinkage in the white population and increase in the other population. The heterogeneity starts to decrease post 2019 when the white population became a minority group and the other population became a majority group. This change shows the importance of the axiom 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 3 shows that when the Asian population and multi-race population is considered separately, heterogeneity actually increased post 2019, as the groups, at a glance, are proportionally growing. Such distributional changes is 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 judgement 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 to why these units needs 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 the version of DUH where $p = 2$, and I would recommend others to do the same.

Figure 2: Comparisons between Different Units for Racial Heterogeneity

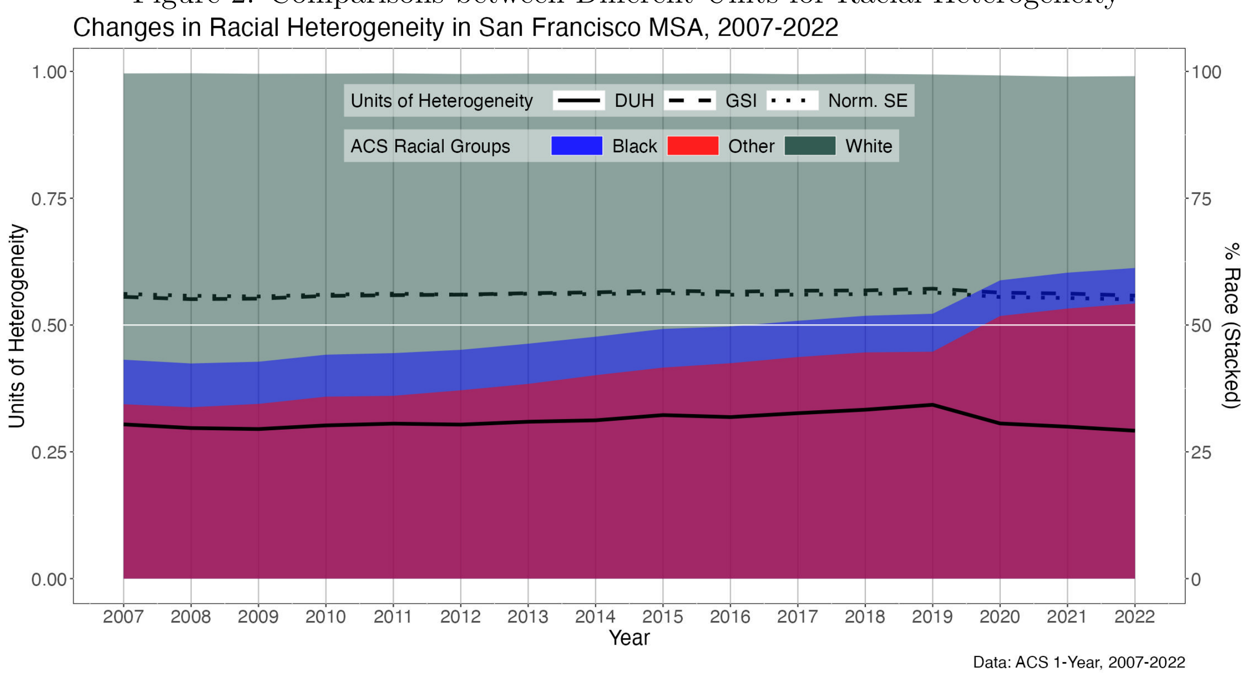
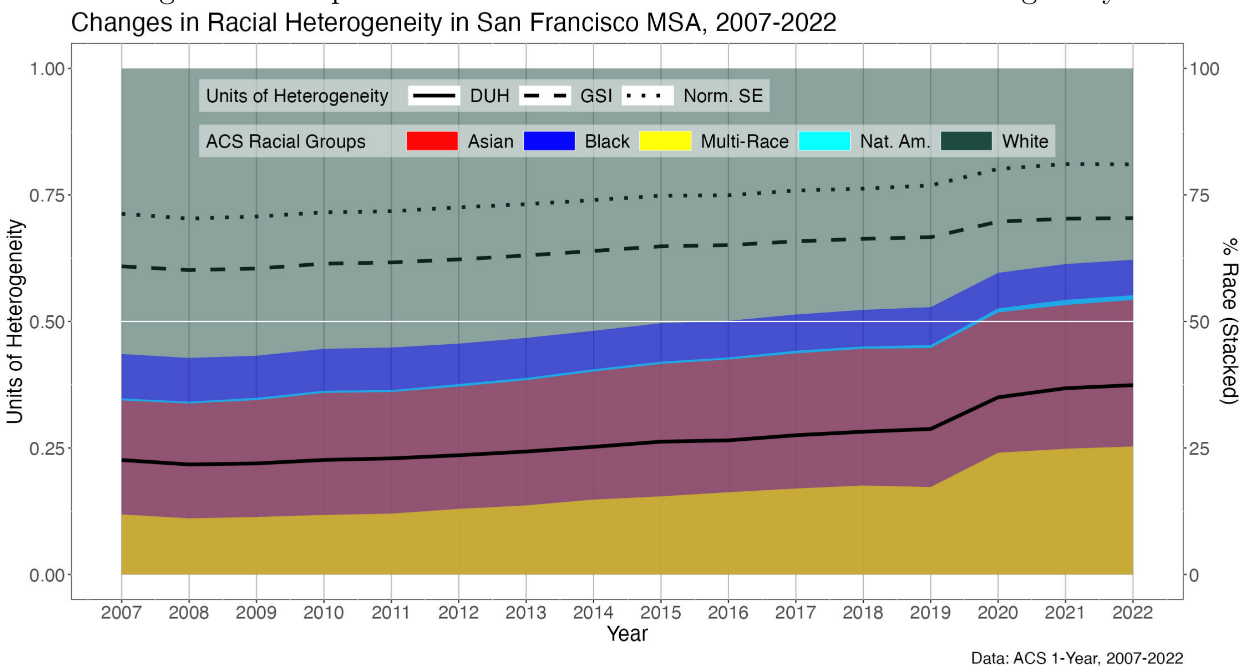


Figure 3: Comparisons between Different Units for Racial Heterogeneity



5.2 Examples

5.2.1 Using DUH for Racial Heterogeneity

Figure 4 shows the progression of racial heterogeneity in San Francisco city proper from 1900-1990[6] using the decennial Census data from IPUMS USA (Ruggles et al. 2024).
From 1980 to 1990, the majority population (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 the decrease in heterogeneity while the larger decrease in SE reflects more of this change in the minority group distribution. DUH on the other hand, follow generally the same trends as GSI and SE, yet it is able to reflect much more of the decrease in evenness in the minority distribution.

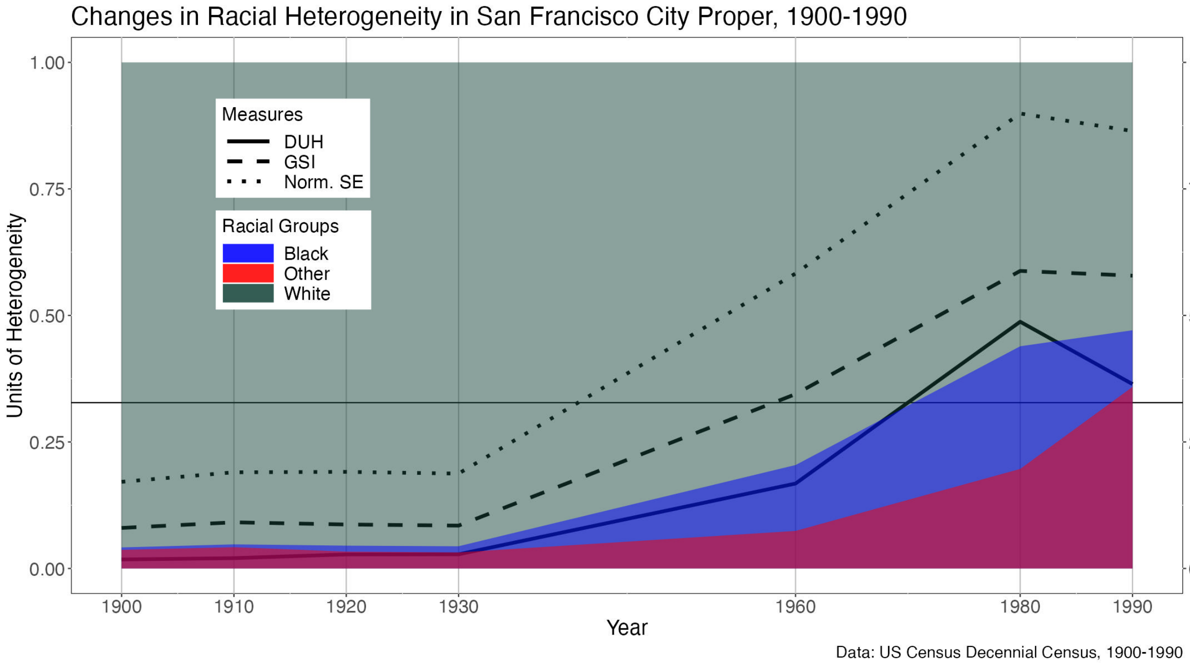
Recall that the main weakness in GSI and SE is that the size of the influence from changes in a group positively correlates with the size of the group. This example shows that DUH is able to dial back the correlation and reflect changes in the evenness of the minority groups.

5.2.2 Using DUH for Product Heterogeneity

Figure 5 illustrates how DUH compares with GSI and DUH in a space that often utilize 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)

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

Figure 4: Comparisons between Different Units for Racial Heterogeneity



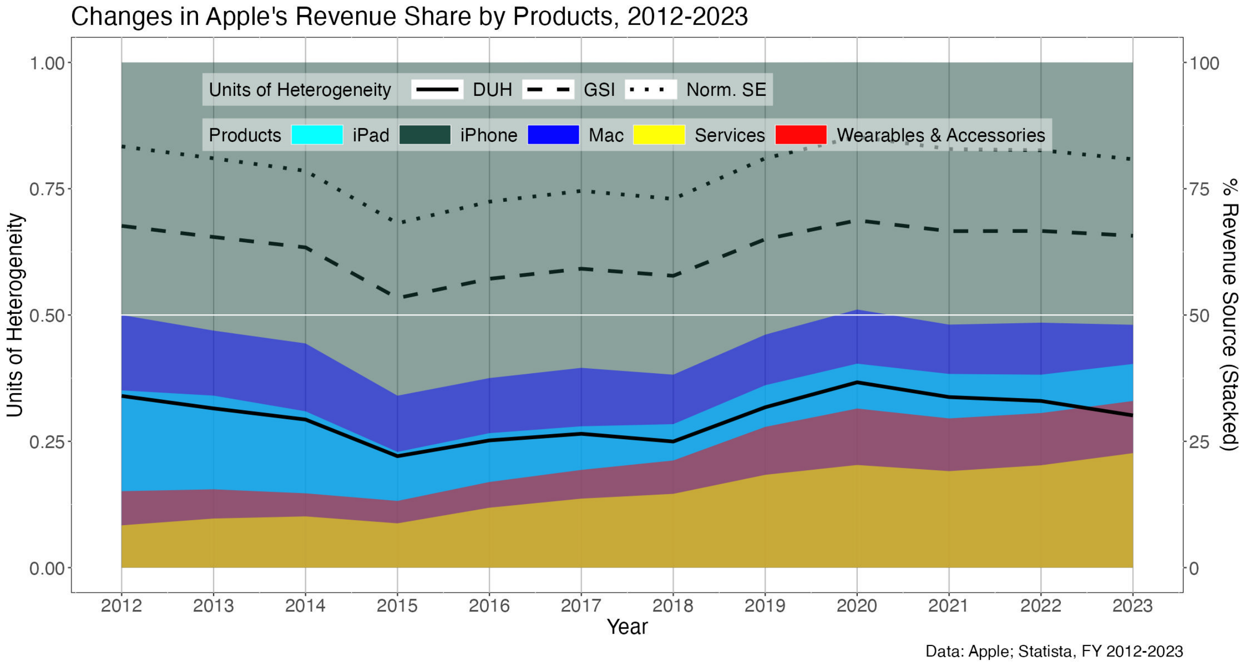
& accessories (like airpods) grew without diminishing the revenue share of iPhones. This decrease in the evenness in minority groups captured by a continuous and sizable decrease in DUH, while decreases in GSI in this period is limited.

6 Summary and Discussion

Building on the axiomatizations of Gini, HHI, and SE, I uniquely characterize the set of units/indices/measures called the *Descriptive Units of Heterogeneity*. This set of units are 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*. Lastly, DUH can likely be extended as a measure for sorting by taking the mean squared differences between the DUH of a system and the DUH of the partitions of said system.

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 in the evolution of heterogeneity in systems.

Figure 5: Comparisons between Different Units for Product Heterogeneity



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. et al. (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.

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

Kolm, S.-C. (1976). Unequal inequalities. ii. *Journal of Economic Theory*, 13(1):82–111.

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

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.

Rothschild, M. and Stiglitz, J. E. (1969). Increasing risk: a definition and its economic consequences. *Cowles Foundation Discussion Paper*.

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.

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.

Appendix: Proofs

Proposition 1: Consider an index $\Phi = \Phi(\varphi, \psi)$ that satisfies *SYM*, *INV*, and *IND*. Holding P_1 constant[7], if ψ is as defined above and Φ is strictly increasing in ψ, 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 $S = (P_1, \dots, P_g, P_{g+1}, \dots, P_G)$ and $S' = (P_1, \dots, P_g - c, P_{g+1} + c, \dots, P_G)$ 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

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

[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 majority group to the minority group, the increase in $\varphi(P_1)$ must dominate the decrease in ψ_p in the case where evenness decreases.

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

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

□

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

Lemma 2: If an index $\Phi(\varphi, \psi)$ of heterogeneity satisfies *Scale Invariance* and *Principle of Proportional Transfers*, it is monotonically decreasing in P_1, and therefore a positive monotonic transformation of $\frac{1}{P_1}$.

Proof:
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 *Scale Invaraince*, *Independence*, *Scale Invariance*, and *Principle of Proportional Transfers*, then φ and ψ must be multiplicatively separable.

Proof:
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 1: If an index $\Phi(\varphi, \psi)$ of heterogeneity satisfies *Independence*, *Scale Invariance*, and *Principle of Proportional Transfers*, 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:
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:

$$\Phi_p(n_1, \ldots, n_G) = -\frac{ln\,(P_1)}{ln(G)}\left[1 - \left(\sum_{g=2}^{G}\left(\tilde{P}_g - \frac{1}{G-1}\right)^p\right)^{\frac{1}{p}}\right]$$
$$\text{where } P_1 = \frac{n_1}{n_1 + \cdots + n_G},\ \tilde{P}_g = \frac{n_g}{n_2 + \cdots + n_g},\ p \in (1, \infty)$$

is the unique class of units that satisfy *Scale Invariance*, *Group Symmetry*, *Independence*, *Principle of Diminishing Transfers*, *Principle of Proportional Transfers*, and *Contractibility.*

Proof:
Notice that to show Φ satisfies *PPT*, 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 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, $\forall x \in \left[0, \frac{(G-1)n_1-\tilde{n}_S}{G}\right]$:

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

$\square$