Barcelona School of Economics

WORKING PAPERS

A Measure of Behavioral Heterogeneity

BSE Working Paper 1373| November 2022

Jose Apesteguia, Miguel Angel Ballester

bse.eu/research

A MEASURE OF BEHAVIORAL HETEROGENEITY

JOSE APESTEGUIA[†] AND MIGUEL A. BALLESTER[‡]

Abstract. In this paper we propose a novel way to measure behavioral heterogeneity in a population of stochastic individuals. Our measure is choice-based; it evaluates the probability that, over a sampled menu, the sampled choices of two sampled individuals differ. We provide axiomatic foundations for this measure, and a decomposition result that separates heterogeneity into its intra- and inter-personal components.

Keywords: Heterogeneity; Intra-personal; Inter-personal; Axiomatic Foundations.
JEL classification numbers: D01.

1. Introduction

A sound measure for quantifying the behavioral heterogeneity of a population is important in economics. For a start, it is an essential tool for empirically explaining underlying driving forces, such as demographics and education. It can also play a role in prediction exercises, where intuition suggests that lower heterogeneity should increase predictive accuracy. In addition, a heterogeneity assessment is an important step in constructing a representative stochastic-agent model capable of capturing the variability within the population. Finally, accounting for heterogeneity may be crucial in guiding welfare analysis.

The behavioral heterogeneity of a population may be the result of two different phenomena. First, the individuals in the population are heterogeneous; that is, they vary in their tastes and, therefore, in their economic choices. Second, the behavior of

Date: December, 2022.

[*]We thank Larbi Alaoui, David Jimenez-Gomez, Andrea Salvanti, Jesse Shapiro and Rafael Suchy for helpful comments. Financial support by FEDER/Ministerio de Ciencia de Innovación (Agencia Estatal de Investigación) through Grant PGC2018-098949-B-I00 and through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), and Balliol College is gratefully acknowledged.

[†]ICREA, Universitat Pompeu Fabra and BSE. E-mail: `jose.apesteguia@upf.edu`.

[‡]University of Oxford. E-mail: `miguel.ballester@economics.ox.ac.uk`.

any given individual is also subject to variation. Making a distinction between these two sources of behavioral heterogeneity, which we refer to as inter-personal and intra-personal, can play an instrumental role in applications. For instance, while classical welfare tools seem appropriate for dealing with heterogeneity driven mainly by inter-personal variability, in the presence of widespread intra-personal heterogeneity, the welfare approach can borrow from the growing literature on behavioral welfare analysis.

In this paper we propose a choice-based measure of behavioral heterogeneity. To allow for the possibility of both inter- and intra-personal variability, we formalize an individual as a random utility model (RUM) and a population as a distribution over RUMs. Since RUMs represent the most standard random choice model, it seems appropriate to use them as the basis for building a measure of behavioral heterogeneity.[1] Thus, we measure behavioral heterogeneity as the probability that, over a sampled menu, the sampled choices of two sampled individuals differ. We call this measure choice heterogeneity, CH. This is an intuitive measure of heterogeneity, which, as we will see in Section 2, sits well with traditional diversity measurement in various fields. In addition, as we will argue in Section 5, it is easily implementable in practice, and thus convenient to use.

In Section 4 we provide axiomatic foundations for CH, which rest on the following ideas. First, CH has the property that two populations with the same representative RUM, i.e. the same convex combination of all individual RUMs in the respective populations, have the same heterogeneity. Second, CH can be decomposed as a weighted sum of the heterogeneity of populations formed by two deterministic individuals. Finally, CH satisfies a simple monotonicity principle by which an increase in choice divergence augments heterogeneity. In order to gain a deeper understanding of the setting and the measure, we first show, both formally and through examples, that CH satisfies these three properties. Then, in Theorem 1, we show that these properties are not only necessary but also sufficient.

Having proposed, and axiomatized, our choice-based measure of behavioral heterogeneity, in Section 5 we elaborate on the decomposition of CH into intra- and inter-personal heterogeneity. For this, we start by arguing that CH can be obtained as the Euclidean proximity between the following two stochastic choice functions: that obtained from the aggregate choices of the population, and that with the highest possible

[1]In Section 7 we argue that our measure of behavioral heterogeneity readily extends to a large class of formalizations of individual random behavior.

variability, which corresponds to uniformly random choices. Then, in Proposition 4, we use this alternative representation of CH to show that our measure can be decomposed into the following two terms: (i) the weighted average of the Euclidean proximity between the stochastic choices of each individual in the population and uniformly random choices and (ii) the weighted average of the Euclidean distance between the stochastic choice functions of each pair of individuals in the population. Thus, term (i) represents intra-personal heterogeneity, and term (ii) inter-personal heterogeneity within the population. Section 6 provides further discussion of these two components of heterogeneity. Comparative statics results are established with respect to intra-personal heterogeneity, while inter-personal heterogeneity is shown to be useful when analyzing a combination of sub-populations.

2. Related Literature

This paper belongs to a long tradition of research in a variety of disciplines such as statistics, linguistics, sociology, quantum mechanics, information theory and economics, where diversity has been measured on the basis of the probability that two random extractions produce different outcomes (see, for example, the measure of diversity of Simpson (1949), the measure of linguistic diversity of Greenberg (1956), the measure of population diversity of Lieberson (1969), the purity parameter in Leonhardt (1997), the residual variance in Ely, Frankel and Kamenica (2015) or its logarithmic version known as the Rényi or collision entropy, and the Herfindahl-Hirschman index of market concentration). Our approach differs in that: (i) we are concerned with choice behavior, which involves a number of overlapping situations (i.e., choices from not just one, but different menus), (ii) we allow for two sources of variability, within and across individuals and, (iii) our treatment is axiomatically founded.

Economics uses a number of approaches for measuring inter-personal variability as it relates to phenomena such as polarization and segregation. Esteban and Ray (1994) measures polarization based on income and wealth distributions, Frankel and Volij (2011) studies school segregation based on between-school distributions, Baldiga and Green (2013) provide a choice-based analysis of consensus, and Gentzkow, Shapiro and Taddy (2019) studies partisanship based on the predictability of party speeches. We differ from the above works in our fundamental interest, which is to capture both intra- and inter-personal behavioral heterogeneity.

There is a large body of applied literature using specific collections of random utility models to describe the behavior of a population. A prominent example is mixed-logit, also known as random-coefficients or random-parameters logit, in which a distribution of individual Luce behaviors is entertained (see Train, 2009).[2] We depart from this literature by offering a measure of heterogeneity.

3. Preliminaries

Consider a finite set of alternatives X. Denote by $\mathcal{A}$ the collection of all subsets of X with at least two alternatives, which we call menus, and by $\mathcal{P}$ the collection of all linear orders over X, which we call preferences. A Random Utility Model (RUM) ψ is a probability distribution on $\mathcal{P}$ interpreted such that, when choosing from menu $A \in \mathcal{A}$, each preference $P \in \mathcal{P}$ is realized with probability $\psi(P)$ and maximized. As a result, RUM choices are stochastic. Denoting by $m(A, P)$ the maximal alternative in menu A according to preference P, and by $\mathbb{I}_{[\cdot]}$ the indicator function which takes the value 1 when the statement in brackets is true and 0 otherwise, the probability that RUM ψ selects alternative a in menu A is equal to:[3]

$$\rho_\psi(a, A) = \sum_P \psi(P) \cdot \mathbb{I}_{[a=m(A,P)]}.$$

We denote by Ψ the set of all RUMs and by Ψ^D the set of all RUMs that are deterministic, i.e., that assign mass 1 to a single preference. For the latter class, we denote by ψ_P the deterministic RUM associated to preference P. In addition, we denote by $\psi_{\mathcal{U}}$ the (uniform) RUM in which all preferences have the same mass.

A population is a probability distribution over the space of RUMs that assigns strictly positive mass to only a finite number of them, i.e., an object with the form

$$\theta = [\theta_1, \theta_2, \dots, \theta_m; \psi_1, \psi_2, \dots, \psi_m].$$

To fix ideas, we can interpret ψ_i as the behavior of a type of individual, with θ_i describing its mass in the population, i.e. $\sum_i \theta_i = 1$. We denote by Θ the set of all populations and by Θ^D the set of all deterministic populations, i.e., those with the form $[\theta_1, \theta_2, \dots, \theta_m; \psi_{P_1}, \psi_{P_2}, \dots, \psi_{P_m}]$, which assign mass only to deterministic RUMs. In words, a deterministic population represents the case of a population in which individuals are deterministic but possibly heterogeneous. Alternatively, denote by Θ^{hom}

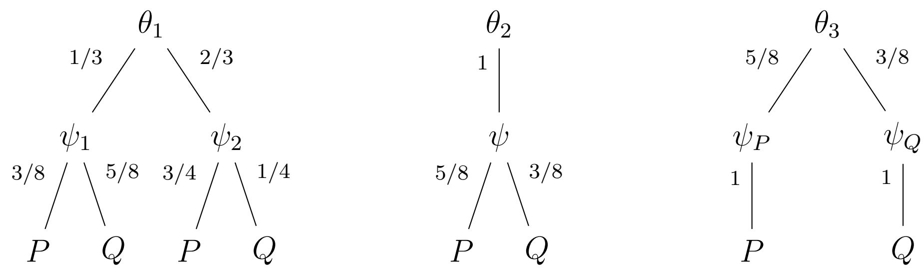
[2]Given the relevance of the Luce and mixed-logit models in applications, we use them to illustrate some of our results.

[3]For ease of exposition, we avoid the specification of any unconstrained domains in the summands.

the set of all populations that are homogeneous, i.e., taking the form $[1; \psi]$. That is, a homogeneous population represents the case of a population in which all individuals are identical to each other, although their behavior possibly admits randomness.

Example 1. Consider the binary set $X = \{x, y\}$. $\mathcal{P}$ contains only two preferences, xPy and yQx and, consequently, any RUM ψ can be identified by the value $\psi(P) \in [0, 1]$ (since $\psi(Q) = 1 - \psi(P)$ is uniquely determined). Let us consider three populations of differing nature, represented graphically in Figure 1.

FIGURE 1. Populations in Example 1.



Population $\theta^1 = [\frac{1}{3}, \frac{2}{3}; \frac{3}{8}, \frac{3}{4}]$ involves two RUMs, determined by the values $\psi_1(P) = \frac{3}{8}$ and $\psi_2(P) = \frac{3}{4}$, with masses $\frac{1}{3}$ and $\frac{2}{3}$ respectively. That is, population θ^1 is neither deterministic nor homogeneous. Population $\theta^2 = [1; \frac{5}{8}]$ is a homogeneous population where all individuals use the non-deterministic RUM that places probability $\frac{5}{8}$ on P. Finally, population $\theta^3 = [\frac{5}{8}, \frac{3}{8}; \psi_P, \psi_Q]$ is a deterministic population involving the two deterministic RUMs, ψ_P and ψ_Q, with masses $\frac{5}{8}$ and $\frac{3}{8}$ respectively. $\square$

4. Behavioral Heterogeneity

We measure heterogeneity as the probability that, over a sampled menu, the sampled choices of two sampled individuals differ. To formalize this notion, consider a distribution λ over $\mathcal{A}$, with $\lambda(A) \geq 0$ describing the probability with which menu A is sampled. Distribution λ may reflect the relative frequency of menus in the dataset, or some judgement by the analyst as to the relative importance of the menus.[4] Formally, the choice heterogeneity of population θ is:

$$\mathsf{CH}_\lambda(\theta) = \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_a \rho_{\psi_i}(a, A)(1 - \rho_{\psi_j}(a, A)).$$

[4]We allow for the possibility that λ assigns zero value to some menus to cover those cases in which the analyst makes no observation on such menus or is not interested in them.

Example 1 (continued). Since there are only two alternatives, it must be that $\lambda(\{x,y\})=1$. Considering population θ^1, we have $\mathsf{CH}_\lambda(\theta^1)=\frac{1}{3}[\frac{1}{3}(\frac{3}{8}\frac{5}{8}+\frac{5}{8}\frac{3}{8})+\frac{2}{3}(\frac{3}{8}\frac{1}{4}+\frac{5}{8}\frac{3}{4})]+\frac{2}{3}[\frac{1}{3}(\frac{3}{4}\frac{5}{8}+\frac{1}{4}\frac{3}{8})+\frac{2}{3}(\frac{3}{4}\frac{1}{4}+\frac{1}{4}\frac{3}{4})]=\frac{15}{32}$. □

We now discuss three plausible properties for a measure of behavioral heterogeneity. To help with the presentation of our characterization result in Section 4.4, we introduce each property in relation to a generic heterogeneity function $\mathsf{H}:\Theta\to\mathbb{R}_+$, which assigns a level of heterogeneity to any possible population, such that $\mathsf{H}(\theta)=0$ if and only if $\theta\in\Theta^D\cap\Theta^{hom}$. Notice that any population in $\Theta^D\cap\Theta^{hom}$ takes the form $[1;\psi_P]$, with all individuals being described by the same, deterministic, behavior. It is apparent that these populations are the only ones in which there is no behavioral variation whatsoever, and hence our basic assumption.

4.1. **Reduction.** The space of RUMs is convex. As a consequence, any population θ admits the construction of an associated RUM, denoted by ψ_θ and called the representative RUM of θ, by using the convex combination of RUMs in the population, with weights equal to their corresponding masses. Formally, the representative RUM of θ is

$$\psi_\theta=\sum_i \theta_i\psi_i.$$

Importantly, the aggregated choices of the homogeneous population $[1;\psi_\theta]$, in which everybody acts according to the representative RUM ψ_θ, are indistinguishable from those of population θ. Therefore, the following is a natural property for a choice-based measure of heterogeneity.[5]

Reduction. $\mathsf{H}(\theta)=\mathsf{H}([1;\psi_\theta])$.

Proposition 1. CH_λ *satisfies Reduction.*[6]

Example 1 (continued). The representative RUM of population θ^1 is $\psi_{\theta^1}(P)=\frac{1}{3}\frac{3}{8}+\frac{2}{3}\frac{3}{4}=\frac{5}{8}$. Hence, the homogeneous population associated to θ^1 is $[1;\psi_{\theta^1}]=\theta^2$. Notice that a direct computation of heterogeneity gives $\mathsf{CH}_\lambda(\theta^2)=\frac{5}{8}\frac{3}{8}+\frac{3}{8}\frac{5}{8}=\frac{15}{32}=\mathsf{CH}_\lambda(\theta^1)$, as claimed by Reduction. It is apparent that the representative RUM of the deterministic

[5]The reader may wish to consider the following analogy with a setting involving lotteries. An individual can be thought of as a simple lottery over preferences, while a population is a compound lottery. Reduction accords equivalent treatment to the compound lottery and to the simple lottery which it induces.

[6]All the proofs are contained in the Appendix.

population θ^3 is the same. Thus, $\mathsf{CH}_\lambda(\theta^3) = \frac{5}{8}[\frac{5}{8}(1\cdot 0 + 0\cdot 1) + \frac{3}{8}(1\cdot 1 + 0\cdot 0)] + \frac{3}{8}[\frac{5}{8}(0\cdot 0 + 1\cdot 1) + \frac{3}{8}(0\cdot 1 + 1\cdot 0)] = \frac{15}{32} = \mathsf{CH}_\lambda(\theta^1) = \mathsf{CH}_\lambda(\theta^2)$ holds. $\square$

The above example illustrates that the same level of heterogeneity can result from intra-personal heterogeneity only, with all individuals being equal to the representative RUM (as in the homogeneous population θ^2), or from inter-personal heterogeneity only, where all individuals are deterministic, with weights given by the distribution of preferences in the representative RUM (as in the deterministic population θ^3), or from a combination of the two (as in population θ^1). The three populations belong to the same iso-heterogeneity level, an equivalence that is implied by Reduction because these three populations have the same representative RUM. In other words, by linking all populations that share a common representative RUM, Reduction delineates the trade-off between the two sources of heterogeneity.

4.2. **Decomposition.** We now consider a deterministic population $\theta \in \Theta^D$, and discuss the possibility of decomposing its heterogeneity as an aggregation of sub-populations. In particular, consider hypothetical sub-populations each formed exclusively by two different deterministic RUMs, with weights in proportion to their masses in the original population, i.e., sub-populations with the form $[\frac{\theta_i}{\theta_i+\theta_j}, \frac{\theta_j}{\theta_i+\theta_j}; \psi_{P_i}, \psi_{P_j}]$.[7] Now, in order to understand the heterogeneity of θ based on that of the binary sub-populations, we should correct back their heterogeneity by the inverse of the normalizing factors, $(\theta_i + \theta_j)^2$. Formally,

Decomposition. For every $\theta \in \Theta^D$, $\mathsf{H}(\theta) = \sum_{i<j} (\theta_i + \theta_j)^2 \ \mathsf{H}([\frac{\theta_i}{\theta_i+\theta_j}, \frac{\theta_j}{\theta_i+\theta_j}; \psi_{P_i}, \psi_{P_j}])$.

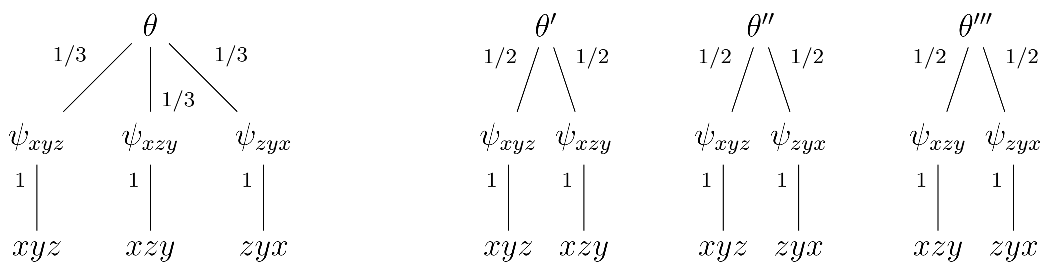
Proposition 2. CH_λ *satisfies Decomposition.*

Example 2.[8] Let $X = \{x, y, z\}$ and the distribution over menus λ. Consider the population $\theta = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}; \psi_{xyz}, \psi_{xzy}, \psi_{zyx}]$, and the subpopulations $\theta' = [\frac{1}{2}, \frac{1}{2}; \psi_{xyz}, \psi_{xzy}]$, $\theta'' = [\frac{1}{2}, \frac{1}{2}; \psi_{xyz}, \psi_{zyx}]$, and $\theta''' = [\frac{1}{2}, \frac{1}{2}; \psi_{xzy}, \psi_{zyx}]$, represented graphically in Figure 2. The heterogeneity of θ is then equal to $\mathsf{CH}_\lambda(\theta) = \lambda(\{x, y\})\frac{1}{9}\cdot 4 + \lambda(\{x, z\})\frac{1}{9}\cdot 4 + \lambda(\{y, z\})\frac{1}{9}\cdot 4 + \lambda(\{x, y, z\})\frac{1}{9}\cdot 4 = \frac{4}{9}$. Decomposition states that we can also see this as $(\frac{1}{3}+$

[7]The RUMs are taken to be different since they are deterministic, and hence there is no role for intra-personal heterogeneity.

[8]We write preferences in the order induced over the alternatives, reading from left to right.

FIGURE 2. Populations in Example 2.



$\frac{1}{3})^2\mathsf{CH}_\lambda([\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{xzy}])+(\frac{1}{3}+\frac{1}{3})^2\mathsf{CH}_\lambda([\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{zyx}])+(\frac{1}{3}+\frac{1}{3})^2\mathsf{CH}_\lambda([\frac{1}{2},\frac{1}{2};\psi_{xzy},\psi_{zyx}]) = \frac{4}{9}\lambda(\{y,z\})\frac{1}{2} + \frac{4}{9}\frac{1}{2} + \frac{4}{9}[\lambda(\{x,y\}) + \lambda(\{x,z\}) + \lambda(\{x,y,z\})]\frac{1}{2} = \frac{4}{9}$. □

4.3. **Monotonicity.** Finally, let us discuss a monotonicity property involving only populations with two equally-likely deterministic RUMs, i.e. with the form $[\frac{1}{2},\frac{1}{2};\psi_P,\psi_Q]$. We will often refer to these populations as couples. In order to state the monotonicity property let us now consider collections of couples $C = \{[\frac{1}{2},\frac{1}{2};\psi_{P^n},\psi_{Q^n}]\}_{n=1}^{N}$. Let us entertain the case in which there are two equally-sized collections of couples C and C' with $N = N'$, and that, whatever the menu at hand, we unequivocally observe a larger number of choice-disagreements in the first of the two. In such a case, it is natural to conclude that the average heterogeneity of this first collection of couples must be larger. Formally, for any C, denote by $\Delta_A(C)$ the number of couples in C for which the two preferences involved disagree over menu A, and by $\overline{\mathsf{H}}(C) = \frac{\sum_n \mathsf{H}([\frac{1}{2},\frac{1}{2};\psi_{P^n},\psi_{Q^n}])}{N}$ the average heterogeneity of all couples in collection C.

Monotonicity. Let C and C' be two equally-sized collections of couples. If $\Delta_A(C) \geq \Delta_A(C')$ for every $A \in \mathcal{A}$, then $\overline{\mathsf{H}}(C) \geq \overline{\mathsf{H}}(C')$.

Proposition 3. CH_λ *satisfies Monotonicity.*

Example 2 (continued). Let $C = \{[\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{xzy}],[\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{zyx}],[\frac{1}{2},\frac{1}{2};\psi_{xzy},\psi_{zyx}]\}$ be the collection of couples related to population θ. If we consider the vector of disagreements $\Delta(\cdot) = (\Delta_{\{x,y\}}(\cdot),\Delta_{\{x,z\}}(\cdot),\Delta_{\{y,z\}}(\cdot),\Delta_{\{x,y,z\}}(\cdot))$, it is immediate that $\Delta(C) = (2,2,2,2)$. Now, let us consider two other, equally-sized, collections of couples. Collection C' is equal to $\{[\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{zxy}],[\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{zyx}],[\frac{1}{2},\frac{1}{2};\psi_{zxy},\psi_{zyx}]\}$, while collection C'' is equal to $\{[\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{yxz}],[\frac{1}{2},\frac{1}{2};\psi_{xyz},\psi_{zyx}],[\frac{1}{2},\frac{1}{2};\psi_{yxz},\psi_{zyx}]\}$. Since $\Delta(C') = (2,2,2,2)$ and $\Delta(C'') = (2,2,2,3)$, Monotonicity implies that the average heterogeneity of couples in C and C' must be equal, and lower than the average heterogeneity of couples in C''. Indeed, our computation above showed that the average heterogeneity of couples in C was $\frac{1}{3}$. Direct computation shows that this is equal to that of C' and below that of C'' which is $\frac{1+\frac{\lambda(\{x,y,z\})}{2}}{3}$. Notice that, using Decomposition, this effectively implies that $\mathsf{CH}_\lambda([\frac{1}{3},\frac{1}{3},\frac{1}{3};\psi_{xyz},\psi_{xzy},\psi_{zyx}]) = \mathsf{CH}_\lambda([\frac{1}{3},\frac{1}{3},\frac{1}{3};\psi_{xyz},\psi_{zxy},\psi_{zyx}]) \leq \mathsf{CH}_\lambda([\frac{1}{3},\frac{1}{3},\frac{1}{3};\psi_{xyz},\psi_{yxz},\psi_{zyx}])$. $\square$

4.4. **A Characterization of CH.** We now show that the three properties discussed above are not only necessary but also sufficient for our measure of heterogeneity.

Theorem 1. H *satisfies Reduction, Decomposition and Monotonicity if and only if there exists a probability distribution* λ *on* $\mathcal{A}$ *and* $k > 0$ *such that* $\mathsf{H} = k \cdot \mathsf{CH}_\lambda$.

Reduction renders the heterogeneity of a population θ equal to that of the homogenous population formed by its representative RUM $[1, \psi_\theta]$. Thus, we consider the deterministic population θ^d that assigns the same probability to every preference as the representative RUM of θ, that is, ψ_θ. Hence, since θ and θ^d have the same representative RUM, Reduction implies that they must have the same heterogeneity. Next, by Decomposition, the heterogeneity of θ^d can be directly broken down into the aggregation of the heterogeneities across sub-populations with the form $[1-\gamma, \gamma; \psi_P, \psi_Q]$, as long as the ratio between $(1-\gamma)$ and γ is equal to the ratio between the masses of preferences P and Q in θ^d. Moreover, we show in the proof that the heterogeneity of population $[1-\gamma, \gamma; \psi_P, \psi_Q]$ can indeed be re-expressed as a product of two terms: (i) a function depending on γ, and (ii) the heterogeneity of the couple involving the same preferences $[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$. This function is actually the logistic map which yields $\mathsf{H}([1-\gamma, \gamma; \psi_P, \psi_Q]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q])$. Thus, we can express the heterogeneity of any population as a weighted additive sum of the heterogeneity of all possible couples, with weights derived from the masses of each preference in the population.

The remaining step in the proof is to obtain the contribution to heterogeneity of each menu A and find the means to link it to the above representation. The difficulty stems from the fact that, generally speaking, it is impossible to find a couple that differs over a single menu A only. Hence, the proof requires the identification of two collections of couples for which the Δ-vectors differ only in menu A, and the application of Monotonicity to these collections. Thus, the difference in heterogeneity between these two collections must correspond to menu A. The proof shows that these added values can be normalized into a probability distribution λ over $\mathcal{A}$ and hence, the heterogeneity of any given population can be expressed as (a scalar transformation of) CH_λ.

5. Alternative Representations of CH

In this section we elaborate on three alternative representations of CH. The first emphasizes the computational convenience of the measure. The second links CH with Euclidean distances in the space of choice functions. The third decomposes CH into its inter- and intra-personal components.

5.1. **A matrix representation of CH.** The proof of Theorem 1 shows that we can evaluate choice heterogeneity by using the frequency of each preference in the representative RUM and the heterogeneity level of every possible couple. Denote by $\mathcal{C}_\lambda$ the $|\mathcal{P}| \times |\mathcal{P}|$-matrix compiling twice the heterogeneity value of each couple. This is a symmetric matrix with zeros in the diagonal and entries for preferences P and Q being equal to the sum of λ-weights of menus where there are choice differences between the two preferences. It is important to stress that this matrix is independent of the specific distribution over RUMs, and hence independent of the population, since it is characterized by the disagreements between preferences, weighted by measure λ. Therefore, the matrix does not need to be recalculated for the analysis of different populations, or for behavioral variations within a population, which is computationally convenient in practice.

Example 2 (continued). Consider the distribution $\bar{\lambda}$ placing equal weight on the four possible menus. Listing the preferences by $xyz, xzy, yxz, yzx, zxy, zyx$, the matrix reporting the heterogeneity of couples is

$$\mathcal{C}_{\bar{\lambda}} = \begin{pmatrix} 0 & 1/4 & 1/2 & 3/4 & 3/4 & 1 \\ 1/4 & 0 & 3/4 & 1 & 1/2 & 3/4 \\ 1/2 & 3/4 & 0 & 1/4 & 1 & 3/4 \\ 3/4 & 1 & 1/4 & 0 & 3/4 & 1/2 \\ 3/4 & 1/2 & 1 & 3/4 & 0 & 1/4 \\ 1 & 3/4 & 3/4 & 1/2 & 1/4 & 0 \end{pmatrix}$$

$\square$

Formally, the following is a direct Corollary of Theorem 1.

Corollary 1. $\mathsf{CH}_\lambda(\theta) = \psi_\theta \mathcal{C}_\lambda \psi_\theta^\top$.

Corollary 1 shows that $\mathcal{C}_\lambda$ is a symmetric positive semi-definite matrix and that the choice heterogeneity of any population can be seen, via the representative RUM, as an inner product.[9]

[9] This is due to the fact that $\mathcal{C}_\lambda$ admits a Cholesky factorization.

Example 3. Here, we consider the mixed-logit model, where a population is formed by a collection of Luce RUMs and their corresponding masses. Given population $[\theta_1, \theta_2, \ldots, \theta_m; \psi_1, \psi_2, \ldots, \psi_m]$ and preference P described by $x_1 P x_2 P \ldots x_{N-1} P x_N$, we can obtain the probability of P in the representative RUM by using the expression $\psi_\theta(P) = \sum_i \theta_i \prod_{j=1}^{N} \frac{u_i(x_j)}{\sum_{k=j}^{N} u_i(x_k)}$, where u_i is the corresponding Luce vector of i.[10] The computation of choice heterogeneity follows easily from the matrix argument in Corollary 1. Consider, e.g., the case of $X = \{x, y, z\}$ and let $\theta = [\frac{4}{11}, \frac{7}{11}; \psi_1, \psi_2]$, where $u_1 = (u_1(x), u_1(y), u_1(z)) = (1/2, 1/3, 1/6)$, and $u_2 = (u_2(x), u_2(y), u_2(z)) = (4/9, 3/9, 2/9)$. The representative RUM is $\psi_\theta = \frac{1}{495}(144, 86, 115, 50, 58, 42)$, and using $\bar{\lambda}$, as in Example 2, $\mathsf{CH}_{\bar{\lambda}}(\theta) = \psi_\theta \mathcal{C}_{\bar{\lambda}} \psi_\theta^\top = .5$. □

5.2. **An Euclidean representation of CH.** We now show that the choice heterogeneity of any population can be seen as a (λ-weighted) Euclidean proximity between the stochastic choice function of the representative RUM and uniformly random behavior.[11] Formally, given any two RUMs ψ and ψ', define the λ-Euclidean distance between their associated stochastic choice functions by

$$d_\lambda(\rho_\psi, \rho_{\psi'}) = \sum_A \lambda(A) \sum_a [\rho_\psi(a, A) - \rho_{\psi'}(a, A)]^2.$$

Consider the constant $\beta_\lambda = \sum_A \lambda(A) \frac{n_A - 1}{n_A}$, where n_A is the number of alternatives in menu A.

Proposition 4. $\mathsf{CH}_\lambda(\theta) = \beta_\lambda - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}}) = \max_{\psi \in \Psi} d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}}) - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}})$.

Proposition 4 first shows that the choice heterogeneity of a population is inversely related to the distance between the stochastic choice function of the representative RUM and uniform choices. Moreover, the second part of Proposition 4 shows that the constant β_λ is in fact the maximum distance between any individual in the population and uniform choices. Hence, the heterogeneity of a population can be understood as

[10] A Luce RUM is usually described by means of a strictly positive real value function u, such that the choice probability of x in menu A is $\frac{u(x)}{\sum_{y \in A} u(y)}$. Without loss of generality, we can normalize u to satisfy $\sum_{x \in X} u(x) = 1$. Hence $u(x)$ can be understood as the probability of choosing x in X. Then, for every menu A, the choice probabilities are simply conditional probabilities. Moreover, notice that a Luce model may admit different RUM representations but, since all of them generate the same stochastic choice function, this is inconsequential for our analysis.

[11] All our analysis uses the square of Euclidean distances. To simplify the presentation, we just write Euclidean all along.

the Euclidean proximity between the aggregate behavior of the population and uniform behavior.

Example 1 (continued). Proposition 4 establishes that $\mathsf{CH}_\lambda(\theta) = \beta_\lambda - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}})$. Since there is only one binary menu, it must be that $\beta_\lambda = \frac{1}{2}$. Now, using our convention to represent RUMs in this simple setting by describing the probability associated with preference P, $\psi_\mathcal{U} = \frac{1}{2}$. Recall that $\psi_{\theta^1} = \frac{5}{8}$ and hence, it must be that $\mathsf{CH}_\lambda(\theta^1) = \frac{15}{32} = \frac{1}{2} - [(\frac{5}{8} - \frac{1}{2})^2 + (\frac{3}{8} - \frac{1}{2})^2]$. □

5.3. **A decomposition of CH into intra- and inter-personal heterogeneity.** We now show that the Euclidean representation of CH in the section above enables us to decompose choice heterogeneity into its intra- and inter-personal components.

Proposition 5. $\mathsf{CH}_\lambda(\theta) = \sum_i \theta_i[\beta_\lambda - d_\lambda(\rho_{\psi_i}, \rho_{\psi_\mathcal{U}})] + \sum_i \theta_i \sum_{i<j} \theta_j \; d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$.

Proposition 5 shows that choice heterogeneity can be decomposed as the aggregation of two different terms. The first of these terms, $\sum_i \theta_i[\beta_\lambda - d_\lambda(\rho_{\psi_i}, \rho_{\psi_\mathcal{U}})]$, evaluates how close each of the individuals in the population is in relation to uniform choices, weighted by their prevalence in the population. The second term, $\sum_i \theta_i \sum_{i<j} \theta_j \; d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$, evaluates the distance between every pair of individuals in the population, again weighted by their prevalence in the population.

Example 1 (continued). Proposition 5 establishes that choice heterogeneity can be obtained as $\sum_i \theta_i[\frac{1}{2} - d_\lambda(\rho_{\psi_i}, \rho_{\psi_\mathcal{U}})] + \sum_i \theta_i \sum_{i<j} \theta_j \; d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$. Direct computation gives $d_\lambda(\rho_{\frac{3}{8}}, \rho_{\frac{1}{2}}) = (\frac{3}{8} - \frac{1}{2})^2 + (\frac{5}{8} - \frac{1}{2})^2 = \frac{1}{32}$, $d_\lambda(\rho_{\frac{3}{4}}, \rho_{\frac{1}{2}}) = (\frac{3}{4} - \frac{1}{2})^2 + (\frac{1}{4} - \frac{1}{2})^2 = \frac{1}{8}$, and $d_\lambda(\rho_{\frac{3}{8}}, \rho_{\frac{3}{4}}) = (\frac{3}{8} - \frac{3}{4})^2 + (\frac{5}{8} - \frac{1}{4})^2 = \frac{9}{32}$, leading to $\frac{1}{3}(\frac{1}{2} - \frac{1}{32}) + \frac{2}{3}(\frac{1}{2} - \frac{1}{8}) + \frac{1}{3}\frac{2}{3}\frac{9}{32} = \frac{15}{32}$. □

6. Intra-personal and Inter-personal heterogeneity

Here, we build on the decomposition result of Proposition 5 to establish some comparative statics results with respect to intra-personal heterogeneity, and show that the consideration of inter-personal heterogeneity proves useful when analyzing the combination of different sub-populations.

6.1. **Intra-personal heterogeneity.** Given an individual ψ, it would be natural to assess its intra-personal heterogeneity. One approach to this would be to use our measure of heterogeneity over the homogeneous population $[1; \psi] \in \Theta^{hom}$. The logic is straightforward; since there is no behavioral variation across individuals in a homogeneous population, heterogeneity must be purely intra-personal. In other words, the intra-personal heterogeneity of a RUM is simply the probability that, over a sampled menu, two sampled choices of individual ψ differ. Moreover, Proposition 5 provides us with another interpretation of this notion of intra-personal heterogeneity; namely, that it corresponds to the λ-Euclidean proximity between individual behavior and uniform choices. That is, we define

$$Intra_\lambda(\psi) = \mathsf{CH}_\lambda([1;\psi]) = \sum_A \lambda(A) \sum_a \rho_\psi(a,A)(1-\rho_\psi(a,A)) = \beta_\lambda - d_\lambda(\rho_\psi, \rho_{\psi_{\mathcal{U}}}).$$

A characterization of this notion of intra-personal heterogeneity follows directly from the characterization in Section 4, by simply considering non-null maps $S : \Psi \to \mathbb{R}_+$, measuring the intra-personal variability of RUMs, and reformulating Decomposition and Monotonicity in terms of RUMs rather than populations.[12]

Corollary 2. *S satisfies Decomposition and Monotonicity if and only if there exists a probability distribution λ on $\mathcal{A}$ and $k > 0$ such that $S = k \cdot \mathrm{Intra}_\lambda$.*

We now investigate further the structure of intra-personal heterogeneity. For this, we use a particular class of RUMs, namely, those satisfying the property that better alternatives are consistently chosen with larger probability, whatever the menu. Formally, for a given $P \in \mathcal{P}$, we say that ψ is P-consistent if xPy and $\{x,y\} \subseteq A$ implies $\rho_\psi(x,A) \geq \rho_\psi(y,A)$. The notion of P-consistency is related to the well-known notion of weak stochastic transitivity. Any P-consistent RUM satisfies weak stochastic transitivity when binary menus are at stake, but it also requires this choice consistency in the remaining menus. A prominent example of such RUMs is the Luce model, as well as many of its generalizations.

Given two P-consistent RUMs, ψ_1 and ψ_2, we say that the latter is a decentralization of the former if there exist $\epsilon > 0$ and preferences Q_1, Q_2 such that: (i) $\psi_2 = \psi_1 - \epsilon\psi_{Q_1} + \epsilon\psi_{Q_2}$ and (ii) Q_2 is further away from P than Q_1 is, i.e., xPy and xQ_2y imply xQ_1y. That is, the second RUM is obtained from the first by shifting mass from preference Q_1 to preference Q_2, which happens to be further from the central preference P. Proposition 6 shows that, in accordance with intuition, this type of shift increases intra-personal heterogeneity. Indeed, the result is also true when sequential changes

[12] In order to write Decomposition in terms of RUMs, one simply needs to consider binary RUMs in which preferences P_i and P_j are entertained with probabilities $\frac{\psi(P_i)}{\psi(P_i)+\psi(P_j)}$ and $\frac{\psi(P_j)}{\psi(P_i)+\psi(P_j)}$. To write Monotonicity in terms of RUMs, the equivalent of a couple population is needed, i.e., one needs to consider RUMs in which only two preferences P_i and P_j are entertained each with $\frac{1}{2}$ probability.

are considered. Formally, we say that ψ_2 is a sequential decentralization of ψ_1 whenever there is a sequence of decentralizations connecting ψ_1 and ψ_2.[13]

Proposition 6. *If ψ_2 is a sequential decentralization of ψ_1, then* $\mathrm{Intra}_\lambda(\psi_2) \geq \mathrm{Intra}_\lambda(\psi_1)$.

Proposition 6 establishes some intuitive comparative statics on intra-personal heterogeneity for transitive RUMs. We now look further into the special case of the Luce model, in which we can conveniently study intra-personal heterogeneity using the monotone likelihood ratio principle.[14]

Proposition 7. *Suppose that $u_1(x_1) \geq \cdots \geq u_1(x_n)$ and $u_2(x_1) \geq \cdots \geq u_2(x_n)$. If $\frac{u_2(x_j)}{u_2(x_i)} \geq \frac{u_1(x_j)}{u_1(x_i)}$ for every $i < j$,* $\mathrm{Intra}_\lambda(\psi_{u_2}) \geq \mathrm{Intra}_\lambda(\psi_{u_1})$.

Proposition 7 considers two Luce-RUMs with the same central preference. By the monotone likelihood ratio, Luce-RUM u_2 places more mass on worse alternatives, and hence Proposition 7 establishes that it must have a larger amount of intra-personal heterogeneity.

Example 3 (continued). Since the monotone likelihood ratio holds for u_1 and u_2, Proposition 7 implies that $Intra_\lambda(\psi_{u_2}) \geq Intra_\lambda(\psi_{u_1})$. Since $\psi_{u_1} = \frac{1}{60}(20, 10, 15, 5, 6, 4)$ and $\psi_{u_2} = \frac{1}{315}(84, 56, 70, 35, 40, 30)$, the matrix computation discussed in Section 5 yields $Intra_\lambda(\psi_{u_1}) = .48$ and $Intra_\lambda(\psi_{u_2}) = .51$. Consider now the representative RUM ψ_θ. Since this is not a Luce RUM, Proposition 7 cannot be applied. However, ψ_θ happens to be a transitive RUM, and it can be seen that ψ_{u_2} is a decentralization of ψ_θ, which in turn is a decentralization of ψ_{u_1}. Hence, Proposition 6 implies that $Intra_\lambda(\psi_\theta) \in [Intra_\lambda(\psi_{u_1}), Intra_\lambda(\psi_{u_2})]$. Notice that we have already computed $Intra_\lambda(\psi_\theta) = \mathsf{CH}_\lambda(\theta) = .50$, as claimed in the result. □

6.2. **Inter-personal heterogeneity.** Proposition 5 provides a decomposition of total heterogeneity into intra-personal and inter-personal components. The inter-personal part, $\sum_i \theta_i \sum_{i<j} \theta_j \; d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$, is a weighted aggregate of the λ-Euclidean distances among individual behaviors in the population. We now show that this value proves useful when studying changes in heterogeneity by mixing two populations. This is the

[13]The result could be formulated alternatively in terms of first-order stochastic dominance over the space of preferences, partially ordered by their distance to the central preference P.

[14]The required notation is given in Example 3.

case because the reasoning in Proposition 5 can be extended to combinations of any two populations θ and θ'.[15]

Corollary 3. *For every* $\alpha \in [0, 1]$,

$$\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)\theta') = \alpha\mathsf{CH}_\lambda(\theta) + (1-\alpha)\mathsf{CH}_\lambda(\theta') + \alpha(1-\alpha)d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_{\theta'}}).$$

Corollary 3 shows that the behavioral heterogeneity of a mixture of sub-populations is the result of: (i) the weighted average of the original choice-based heterogeneities and (ii) the inter-personal heterogeneity arising from the, possibly different, representative RUMs of each sup-population. The result describes the practical nature of the choice heterogeneity measure when considering existing information on sub-populations. The aggregate heterogeneity can be computed merely from the heterogeneity of the sub-populations and the added inter-population heterogeneity, via the representative agents of these populations. It is thus apparent how heterogeneity responds to some specific aggregations. For example, consider the case in which the two sub-populations have the same heterogeneity. If the sub-populations are not identical, one would expect the level of heterogeneity to increase when the two are combined. Corollary 3 confirms this by showing that the additional heterogeneity can be obtained simply by inspecting the distance between the representative RUMs. Another particular case of interest is that of the tremble model, where a population θ is mixed with a uniform distribution over preferences. Here, since the heterogeneity of uniform choices is higher than that of any other population, the mixing with the uniform distribution produces an increase (through both channels (i) and (ii)) of heterogeneity; the mixture is unequivocally more heterogeneous than the original population θ. In particular,

Corollary 4. *For every* $\alpha \in [0, 1]$, $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1; \psi_\mathcal{U}]) = \beta_\lambda - \alpha^2 d_\lambda(\psi_\theta, \psi_\mathcal{U})$.

Example 1 (continued). Let θ' be the population obtained by mixing α of the original population θ^1 and $1-\alpha$ of uniform behavior, i.e., $\theta' = \alpha\theta^1 + (1-\alpha)[1; \psi_\mathcal{U}] = [\frac{\alpha}{3}, \frac{2\alpha}{3}, 1-\alpha; \frac{3}{8}, \frac{3}{4}, \frac{1}{2}]$. Corollary 3 allows the computation of the heterogeneity of the tremble mixture as $\alpha\frac{15}{32} + (1-\alpha)\frac{1}{2} + \alpha(1-\alpha)\frac{1}{32}$ which, as claimed by Corollary 4, is $\frac{1}{2} - \alpha^2\frac{1}{32}$, a value that increases with the trembling weight $1-\alpha$. $\square$

[15] We write $\alpha\theta+(1-\alpha)\theta'$ to represent the population induced by the combination of sub-populations θ and θ' with weights α and $1-\alpha$.

7. Discussion

Based on the prevalence of RUMs in the modeling of heterogeneity, we have offered a choice-based measure of heterogeneity for populations composed of individuals behaving á la RUM. Notice that our measure of heterogeneity is directly applicable in settings where behavioral structures other than RUMs are in place. In particular, if the individuals in a population can be described by any sort of stochastic choice function, the measure CH_λ is well-defined, and the decomposition into intra-personal and inter-personal heterogeneity described in Proposition 5 holds. Moreover, our characterization result goes through as long as the setting satisfies the following two properties: (i) the domain of individual behaviors must be convex, allowing for the existence of a representative behavior in any population, and (ii) it should be possible to link any menu to a pair of deterministic behaviors, or, possibly, to a collection of pairs of deterministic behaviors, as explained in the discussion after Theorem 1. A simple, general example that meets these two properties is the space of all stochastic choice functions, where no rationality requirement whatsoever is imposed on individuals. This domain is clearly convex and, for any given menu, one can easily construct a pair of deterministic choice functions that differ only over the given menu. Hence, our characterization result can be adapted to this setting.

We close by commenting on the empirical implementation of our measure of choice heterogeneity. The natural dataset would involve multiple choices by different individuals, or different types of individuals, such as those given by age groups, gender, etc. Practitioners would then proceed by estimating the individual RUMs, or, based on the above discussion, by using a preferred stochastic behavioral model. There is a series of papers proposing statistical tests and estimation techniques for a variety of stochastic models that could be used to determine the appropriate class of individual stochastic models and their specification (see, e.g., Halevy, Persitz, and Zrill (2018), Kitamura and Stoye (2018), Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), Barseghyan, Molinari, and Thirkettle (2021), Apesteguia and Ballester (2021), Dardanoni, Manzini, Mariotti, Petri, and Tyson (2022), and de Clippel and Rozen (2022)). Once the individual stochastic models are specified, the application of our measure is direct, as discussed in the main text (see, in particular, Section 5).

Appendix A. Proofs

Proof of Proposition 1: The choice-based heterogeneity of population θ can be rewritten as:

$$\mathsf{CH}_\lambda(\theta) = \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_a \rho_{\psi_i}(a, A)(1 - \rho_{\psi_j}(a, A))$$

$$= \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_P \psi_i(P) \sum_Q \psi_j(Q) \cdot \mathbb{I}_{[m(A,P) \neq m(A,Q)]}$$

$$= \sum_A \lambda(A) \sum_i \sum_P \theta_i \psi_i(P) \sum_j \sum_Q \theta_j \psi_j(Q) \cdot \mathbb{I}_{[m(A,P) \neq m(A,Q)]}$$

$$= \sum_A \lambda(A) \sum_P \psi_\theta(P) \sum_Q \psi_\theta(Q) \cdot \mathbb{I}_{[m(A,P) \neq m(A,Q)]}$$

$$= \sum_A \lambda(A) \sum_a \rho_{\psi_\theta}(a, A)(1 - \rho_{\psi_\theta}(a, A)) = \mathsf{CH}_\lambda([1; \psi_\theta]).$$

■

Proof of Proposition 2: Let $\theta = [\theta_1, \theta_2, \ldots, \theta_m; \psi_{P_1}, \psi_{P_2}, \ldots, \psi_{P_m}]$ be a deterministic population. The probability that a deterministic RUM makes two different choices is zero, and hence the heterogeneity of θ can be written as

$$\mathsf{CH}_\lambda(\theta) = \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_a \rho_{\psi_{P_i}}(a, A)(1 - \rho_{\psi_{P_j}}(a, A)) =$$

$$\sum_A \lambda(A) \sum_{i<j} 2\, \theta_i \theta_j \sum_a \rho_{\psi_{P_i}}(a, A)(1 - \rho_{\psi_{P_j}}(a, A)) = \sum_A \lambda(A) \sum_{i<j} 2\theta_i \theta_j \mathbb{I}_{[m(A,P_i) \neq m(A,P_j)]} =$$

$$\sum_{i<j} (\theta_i + \theta_j)^2 \sum_A \lambda(A) \frac{2\theta_i \theta_j}{(\theta_i + \theta_j)^2} \mathbb{I}_{[m(A,P_i) \neq m(A,P_j)]}$$

$$= \sum_{i<j} (\theta_i + \theta_j)^2\, \mathsf{CH}_\lambda([\frac{\theta_i}{\theta_i + \theta_j}, \frac{\theta_j}{\theta_i + \theta_j}; \psi_{P_i}, \psi_{P_j}]).$$

■

Proof of Proposition 3: The average heterogeneity of the collection of couples $C = \{[\frac{1}{2}, \frac{1}{2}; \psi_{P^n}, \psi_{Q^n}]\}_{n=1}^N$ is:

$$\overline{\mathsf{CH}}_\lambda(C) = \frac{1}{N} \sum_n \sum_A \lambda(A) \frac{1}{2} \cdot \mathbb{I}_{[m(A,P^n) \neq m(A,Q^n)]} = \frac{1}{2N} \sum_A \lambda(A) \sum_n \mathbb{I}_{[m(A,P^n) \neq m(A,Q^n)]}$$

$$= \frac{1}{2N} \sum_A \lambda(A) \Delta_A(C).$$

Given that λ is a positive-valued function, the statement follows. ■

Proof of Theorem 1: Propositions 1 to 3 show the necessity of the axioms. It is also immediate that $\mathsf{CH}_\lambda(\theta) = 0$ if and only if $\theta \in \Theta^D \cap \Theta^{hom}$, as required by our basic assumption over the heterogeneity map. We now prove the sufficiency of all these properties. Let us consider any menu $A \in \mathcal{A}$ and proceed by fixing one pair of different alternatives $\{a, b\} \subseteq A$. Then, for every menu B with the property $\{a, b\} \subseteq B \subseteq A$, let us fix a preference P_B^A satisfying $(X \setminus B)\, P_B\, a\, P_B\, b\, P_B\, (B \setminus \{a, b\})$. By considering the couple formed by preference P_B^A and the preference Q_B^A that is obtained by swapping the position of alternatives a and b in the preference, we are able to define the value

$$\sum_{B:\{a,b\}\subseteq B\subseteq A} (-1)^{|A|-|B|}\ \mathsf{H}\big(\big[\tfrac{1}{2}, \tfrac{1}{2}; \psi_{P_B^A}, \psi_{Q_B^A}\big]\big). \tag{1}$$

Claim 1. Expression (1) is independent of the selected pair of alternatives and collection of preferences. Accordingly, we denote the value defined by expression (1) as $\tau(A)$.

To prove Claim 1, let us fix a menu A and consider any two pairs of different alternatives $\{a, b\}$ and $\{a', b'\}$ in this menu and any two associated collections of preferences $\{P_B^A, Q_B^A\}_{B:\{a,b\}\subseteq B\subseteq A}$ and $\{P'^A_{B'}, Q'^A_{B'}\}_{B':\{a',b'\}\subseteq B'\subseteq A}$. Let us then distinguish the following collections of couples (i) C_1^A is formed by all couples $[\frac{1}{2}, \frac{1}{2}; \psi_{P_B^A}, \psi_{Q_B^A}]$ where $\{a, b\} \subseteq B \subseteq A$ is such that $(-1)^{|A|-|B|} = 1$, (ii) C_2^A is formed by all couples $[\frac{1}{2}, \frac{1}{2}; \psi_{P_B^A}, \psi_{Q_B^A}]$ where $\{a, b\} \subseteq B \subseteq A$ is such that $(-1)^{|A|-|B|} = -1$, (iii) $C_1'^A$ is the collection of all couples $[\frac{1}{2}, \frac{1}{2}; \psi_{P'^A_{B'}}, \psi_{Q'^A_{B'}}]$ where $\{a', b'\} \subseteq B' \subseteq A$ satisfies $(-1)^{|A|-|B'|} = 1$ and, finally (iv) $C_2'^A$ is formed by all couples $[\frac{1}{2}, \frac{1}{2}; \psi_{P'^A_{B'}}, \psi_{Q'^A_{B'}}]$ where $\{a', b'\} \subseteq B' \subseteq A$ is such that $(-1)^{|A|-|B'|} = -1$. It is immediate to see that, for every $S \neq A$, $\Delta_S(C_1^A) = \Delta_S(C_2^A)$ and $\Delta_S(C_1'^A) = \Delta_S(C_2'^A)$, while $\Delta_A(C_1^A) = \Delta_A(C_1'^A) = 1 > 0 = \Delta_A(C_2^A) = \Delta_A(C_2'^A)$. Hence, the Δ-values of the collections of couples $C_1^A \cup C_2'^A$ and $C_2^A \cup C_1'^A$ must coincide and, since they are equally-sized, Monotonicity guarantees that $\sum_{\theta\in C_1^A} \mathsf{H}(\theta) + \sum_{\theta\in C_2'^A} \mathsf{H}(\theta)$ is equal to $\sum_{\theta\in C_2^A} \mathsf{H}(\theta) + \sum_{\theta\in C_1'^A} \mathsf{H}(\theta)$. By rearranging, we obtain

$$\sum_{B:\{a,b\}\subseteq B\subseteq A} (-1)^{|A|-|B|}\ \mathsf{H}\big(\big[\tfrac{1}{2}, \tfrac{1}{2}; \psi_{P_B^A}, \psi_{Q_B^A}\big]\big) = \sum_{\theta\in C_1^A} \mathsf{H}(\theta) - \sum_{\theta\in C_2^A} \mathsf{H}(\theta) =$$

$$\sum_{\theta\in C_1'^A} \mathsf{H}(\theta) - \sum_{\theta\in C_2'^A} \mathsf{H}(\theta) = \sum_{B':\{a',b'\}\subseteq B'\subseteq A} (-1)^{|A|-|B'|}\ \mathsf{H}\big(\big[\tfrac{1}{2}, \tfrac{1}{2}, \psi_{P'^A_{B'}}, \psi_{Q'^A_{B'}}\big]\big) = \tau(A).$$

Claim 2. For every pair of preferences $P, Q \in \mathcal{P}$, it must be that

$$\mathsf{H}([\tfrac{1}{2}, \tfrac{1}{2}, \psi_P, \psi_Q]) = \sum_A \tau(A) \cdot \mathbb{I}_{[m(A,P) \neq m(A,Q)]}.$$

If P is equal to Q, we know by assumption that $\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]) = 0$, as desired. Then, let us assume that $\{A : m(A,P) \neq m(A,Q)\}$ is non-empty, and denote by $n \geq 0$ the number of menus with two alternatives over which P and Q differ. For every menu A such that $m(A,P) \neq m(A,Q)$, denote by C_1^A and C_2^A the corresponding collections of couples defined in the proof of Claim 1.

Consider the two collections of symmetric binary populations: (i) $\bigcup_{A:m(A,P)\neq m(A,Q)} C_1^A$ and (ii) $\bigcup_{A:m(A,P)\neq m(A,Q)} C_2^A \cup \{[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]\}$. Notice that, for every binary menu such that $m(A,P) \neq m(A,Q)$, (i) contains one couple while (ii) contains none. In addition, (ii) has the extra population defined by $[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$. Hence, if $n = 0$, select any preference R and add the population $[1; \psi_R] = [\frac{1}{2}, \frac{1}{2}; \psi_R, \psi_R]$ to (i). If $n > 1$, add $n-1$ copies of the population $[1; \psi_R] = [\frac{1}{2}, \frac{1}{2}; \psi_R, \psi_R]$ to (ii). In any case, we have defined two equally-sized collections of couples which we call, respectively, C and C'.

From the analysis in Claim 1, we know that $\Delta_S(C_1^A) = \Delta_S(C_2^A)$ for every $S \neq A$ and $\Delta_A(C_1^A) = 1 > 0 = \Delta_A(C_2^A)$. Since populations $[\frac{1}{2}, \frac{1}{2}; \psi_R, \psi_R]$ are irrelevant in this respect, and population $[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$ is such that $\Delta_A(\{[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]\}) = 1$ if and only if $m(A,P) \neq m(A,Q)$, it is indeed the case that C and C' have the same vector Δ over all menus. From this it is immediate that $\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_R, \psi_R]) = 0$ and we can then apply Monotonicity to obtain

$$\sum_{A:m(A,P_1)\neq m(A,P_2)} \sum_{\theta \in C_1^A} \mathsf{H}(\theta) = \sum_{A:m(A,P_1)\neq m(A,P_2)} \sum_{\theta \in C_2^A} \mathsf{H}(\theta) + \mathsf{H}([\tfrac{1}{2}, \tfrac{1}{2}; \psi_P, \psi_Q]).$$

It then follows that

$$\mathsf{H}([\tfrac{1}{2}, \tfrac{1}{2}; \psi_P, \psi_Q]) = \sum_{A:m(A,P)\neq m(A,Q)} \Big(\sum_{\theta \in C_1^A} \mathsf{H}(\theta) - \sum_{\theta \in C_2^A} \mathsf{H}(\theta)\Big) = \sum_{A:m(A,P)\neq m(A,Q)} \tau(A).$$

Claim 3. The map λ given by $\lambda(A) = \frac{\tau(A)}{\sum_A \tau(A)}$ is a probability distribution over $\mathcal{A}$.

Given our choice of normalization method, we simply need to show that τ is positive and non-null. To prove positivity, consider any menu A and the corresponding collections C_1^A and C_2^A, as defined in the proof of Claim 1. We know that $\tau(A) = \sum_{\theta \in C_1^A} \mathsf{H}(\theta) - \sum_{\theta \in C_2^A} \mathsf{H}(\theta)$. Hence, if $|A| = 2$, collection C_1^A is formed by a unique population, while collection C_2^A is empty and the positivity of H guarantees the positivity of $\tau(A)$. If $|A| > 2$, collections C_1^A and C_2^A are equally-sized, $\Delta_S(C_1^A) = \Delta_S(C_2^A)$ holds for every $S \neq A$, and $\Delta_A(C_1^A) = 1 > 0 = \Delta_A(C_2^A)$, and again positivity holds. To prove that τ is non-null, assume, by contradiction, that this is not the case. Then, Claim 2 implies that every couple has zero heterogeneity. Since there are couples not belonging to Θ^{hom}, this is a contradiction. Hence, τ must be non-null and λ must be a probability distribution over menus.

Claim 4. For every pair of preferences $P, Q \in \mathcal{P}$ and constant $\gamma \in [0,1]$, it is the case that $\mathsf{H}([1-\gamma, \gamma; \psi_P, \psi_Q]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q])$.

To see this, fix two preferences $P, Q \in \mathcal{P}$. Then consider any two values $\alpha, \beta \in [0,1]$ and the mixing of populations $[1-\alpha, \alpha; \psi_P, \psi_Q]$ and $[1-\beta, \beta; \psi_P, \psi_Q]$ with weights $\frac{\beta}{\alpha+\beta}$ and $\frac{\alpha}{\alpha+\beta}$. That is, let $\theta' = [\frac{\beta}{\alpha+\beta}(1-\alpha), \frac{\alpha}{\alpha+\beta}(1-\beta), \frac{\beta}{\alpha+\beta}\alpha, \frac{\alpha}{\alpha+\beta}\beta; \psi_P, \psi_P, \psi_Q, \psi_Q]$. Since this population is deterministic, the application of Decomposition, together with the fact that homogeneous and deterministic populations have zero heterogeneity, leads to

$$\mathsf{H}(\theta') = 2[(\frac{\beta}{\alpha+\beta})^2\, \mathsf{H}([1-\alpha, \alpha; \psi_P, \psi_Q]) + (\frac{\alpha}{\alpha+\beta})^2\, \mathsf{H}([1-\beta, \beta; \psi_P, \psi_Q])].$$

Since we have $\frac{\beta}{\alpha+\beta}(1-\alpha) + \frac{\alpha}{\alpha+\beta}(1-\beta) = \frac{\alpha+\beta-2\alpha\beta}{\alpha+\beta}$, Reduction guarantees that the heterogeneity of population $[\frac{\alpha+\beta-2\alpha\beta}{\alpha+\beta}, \frac{2\alpha\beta}{\alpha+\beta}; \psi_P, \psi_Q]$ must be equivalent to that of θ', leading to

$$\mathsf{H}([\frac{\alpha+\beta-2\alpha\beta}{\alpha+\beta}, \frac{2\alpha\beta}{\alpha+\beta}; \psi_P, \psi_Q]) =$$

$$2[(\frac{\beta}{\alpha+\beta})^2\mathsf{H}([1-\alpha, \alpha; \psi_P, \psi_Q]) + (\frac{\alpha}{\alpha+\beta})^2\mathsf{H}([1-\beta, \beta; \psi_P, \psi_Q])].$$

Direct manipulation shows that $\mathsf{H}([1-\gamma, \gamma; \psi_P, \psi_Q]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$ must hold.

Claim 5. For every $\theta \in \Theta^D$, $\mathsf{H}(\theta) = \sum_{i<j} 4\,\theta_i\theta_j \mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_{P_i}, \psi_{P_j}])$.

Consider $\theta \in \Theta^D$. The result follows from combining Decomposition and Claim 4.

$$\mathsf{H}(\theta) = \sum_{i<j} (\theta_i + \theta_j)^2\ \mathsf{H}([\frac{\theta_i}{\theta_i+\theta_j}, \frac{\theta_j}{\theta_i+\theta_j}; \psi_{P_i}, \psi_{P_j}])$$

$$= \sum_{i<j} (\theta_i+\theta_j)^2\, 4\frac{\theta_i}{\theta_i+\theta_j}\frac{\theta_j}{\theta_i+\theta_j}\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_{P_i}, \psi_{P_j}]) = \sum_{i<j} 4\,\theta_i\theta_j\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_{P_i}, \psi_{P_j}]).$$

Claim 6. $\mathsf{H} = k \cdot \mathsf{CH}_\lambda$ for some $k > 0$.

Consider any population θ. Construct the unique deterministic population $\theta^d \in \Theta^D$ with the property that, for every $P \in \mathcal{P}$, $\theta^d(\psi_P) = \psi_\theta(P)$ (where, recall that ψ_θ is the representative RUM of θ). From Claim 5, $\mathsf{H}(\theta^d) = \displaystyle\sum_{i<j} 4\,\theta_i^d \theta_j^d \mathsf{H}([\tfrac{1}{2}, \tfrac{1}{2}; \psi_{P_i}, \psi_{P_j}])$. Using Claim 2, we have $\mathsf{H}(\theta^d) = \displaystyle\sum_{i<j} 4\,\theta_i^d \theta_j^d \sum_{A: m(A,P_i) \neq m(A,P_j)} \tau(A)$. We can rewrite this expression as $\mathsf{H}(\theta^d) = k \sum_A \lambda(A) \sum_i \theta_i^d \sum_j \theta_j^d \mathbb{I}_{[m(A,P_i) \neq m(A,P_j)]}$, which, given the fact that θ^d is deterministic, coincides with $\mathsf{CH}_\lambda(\theta^d)$. Now, simply notice that the representative RUM of θ^d coincides with that of θ, and Reduction (and the fact that CH_λ satisfies this property) guarantees that $\mathsf{H}(\theta) = \mathsf{H}(\theta^d) = \mathsf{CH}_\lambda(\theta^d) = \mathsf{CH}_\lambda(\theta)$. This concludes the proof. ■

Proof of Proposition 4: We start by proving a series of useful claims. The first is that, conditional on having sampled the ordered pair of RUMs (ψ, ψ'), the probability that a random choice from ψ disagrees with a random choice from ψ', over a random menu, can be written as:

$$\frac{1}{2}[\mathsf{CH}_\lambda([1;\psi]) + \mathsf{CH}_\lambda([1;\psi']) + d_\lambda(\rho_\psi, \rho_{\psi'})].$$

We call this probability the conditional heterogeneity of (ψ, ψ').

To prove the claim, suppose that we have sampled the ordered pair of RUMs (ψ, ψ'). Conditional heterogeneity is $\sum_A \lambda(A) \sum_a \rho_\psi(a, A)(1 - \rho_{\psi'}(a, A))$, or equivalently

$$\sum_A \lambda(A) \sum_a [\rho_\psi(a,A)(1 - \rho_\psi(a,A)) + \rho_\psi(a,A)(\rho_\psi(a,A) - \rho_{\psi'}(a,A))].$$

By similar reasoning, conditional heterogeneity is also equal to

$$\sum_A \lambda(A) \sum_a [\rho_{\psi'}(a,A)(1 - \rho_{\psi'}(a,A)) + \rho_{\psi'}(a,A)(\rho_{\psi'}(a,A) - \rho_\psi(a,A))].$$

Thus, conditional heterogeneity must be equal to the average of the last two expressions, which is simply

$$\frac{1}{2} \sum_A \lambda(A) \sum_a [\rho_\psi(a,A)(1 - \rho_\psi(a,A)) + \rho_{\psi'}(a,A)(1 - \rho_{\psi'}(a,A))$$

$$+(\rho_\psi(a,A) - \rho_{\psi'}(a,A)^2)] = \frac{1}{2}[\mathsf{CH}_\lambda([1;\psi]) + \mathsf{CH}_\lambda([1;\psi']) + d_\lambda(\rho_\psi, \rho_{\psi'})],$$

as claimed.

Second, we claim that for every population $\theta \in \Theta$, $\mathsf{CH}_\lambda(\theta) = \sum_i \theta_i \mathsf{CH}_\lambda([1;\psi_i]) + \sum_i \theta_i \sum_{i<j} \theta_j\, d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$. To see this, notice that $\mathsf{CH}_\lambda(\theta)$ is simply the aggregation

of conditional heterogeneities across all possible ordered pairs of RUMs weighted by their corresponding sampling probabilities. Hence, we proceed by aggregating the expression given above. Since every RUM ψ_i appears as the first RUM in the sampling with probability θ_i and again, as the second RUM in the sampling with probability θ_i, the aggregation of conditional heterogeneities creates the value $\sum_i \theta_i \mathsf{CH}_\lambda([1;\psi_i])$. Given ψ_i and ψ_j, with $i<j$, these two RUMs appear in the sampling with probability $2\,\theta_i\theta_j$ and given the symmetry of d_λ, the aggregation of all expressions creates the value $\sum_i \theta_i \sum_{i<j} \theta_j \; d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$, thus proving the claim.

Third, we claim that for any RUM ψ, $\mathsf{CH}_\lambda([1;\psi]) = \beta_\lambda - d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}})$ holds. To see this, consider the couple $\theta = [\frac{1}{2}, \frac{1}{2}; \psi, \psi_\mathcal{U}]$. From the previous claim, $\mathsf{CH}_\lambda(\theta) = \frac{1}{2}\mathsf{CH}_\lambda([1;\psi]) + \frac{1}{2}\mathsf{CH}_\lambda([1;\psi_\mathcal{U}]) + \frac{1}{4} d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}})$. Now, notice that, since one of the RUMs involved is uniform, direct computation of the heterogeneity of θ yields $\mathsf{CH}_\lambda(\theta) = \frac{1}{4}\mathsf{CH}_\lambda([1;\psi]) + \frac{3}{4}\beta_\lambda$. By putting these two expressions together, we obtain:

$$\mathsf{CH}_\lambda([1;\psi]) = 3\beta_\lambda - 2\mathsf{CH}_\lambda([1;\psi_\mathcal{U}]) - d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}})$$

$$= 3\beta_\lambda - 2\beta_\lambda - d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}}) = \beta_\lambda - d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}}),$$

which proves the claim.

Now, to prove the statement, note that Proposition 1 guarantees that $\mathsf{CH}_\lambda(\theta) = \mathsf{CH}_\lambda([1;\psi_\theta])$, and by the third claim $\mathsf{CH}_\lambda(\theta) = \beta_\lambda - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}})$ holds. Finally, notice that $\max_{\psi\in\Psi} d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}})$ will be achieved by any RUM belonging to Θ^D, leading to $\sum_A \lambda(A)[(1-\frac{1}{n_A})^2 + (n_A-1)(\frac{1}{n_A}-0)^2] = \sum_A \lambda(A)[\frac{(n_A-1)^2}{n_A^2} + \frac{n_A-1}{n_A^2}] = \sum_A \lambda(A)\frac{n_A-1}{n_A} = \beta_\lambda$, which concludes the proof. ■

Proof of Proposition 5: The proof follows directly from the second and third claims in the proof of Proposition 4. ■

Proof of Proposition 6: Suppose that ψ_2 is a sequential decentralization of ψ_1. By definition, there exists a sequence $\{\psi^j\}_{j=1}^J$ of RUMs such that $\psi^1 = \psi_1$ and $\psi^J = \psi_2$, and ψ^j is a decentralization of ψ^{j-1} for $j = 2, \dots, J$, with the central preference denoted as P. At each stage j, mass $\epsilon_j > 0$ shifts from preference Q_1^j to another preference Q_2^j, i.e., $\psi^{j+1} = \psi^j - \epsilon^j \psi_{Q_1^j} + \epsilon^j \psi_{Q_2^j}$. Since every decentralization can indeed be obtained as a sequence of decentralizations in which the two preferences differ in their ranking of two alternatives, we assume w.l.o.g. that Q_1^j and Q_2^j differ in their ranking of only two alternatives, with $x^j P y^j, x^j Q_1^j y^j$ and $y^j Q_2^j x^j$.

First, consider any menu A that fails to contain either x^j or y^j or such that $m(A, Q_1^j) \neq x_j$. Preferences Q_1^j and Q_2^j have the same maximizer over such a menu and hence, it is evident that $\rho_{\psi^{j+1}}(\cdot, A) = \rho_{\psi^j}(\cdot, A)$, i.e., the transfer of mass is irrelevant for the intra-personal heterogeneity over such menus. Second, consider any menu satisfying $\{x^j, y^j\} \subseteq A$ and $x^j = m(A, Q_1^j)$. Within such menus, the transfer of mass increases the choice probability of alternative y^j while reducing that of alternative x^j, with no other changes for the remaining alternatives. Thus, we know that $\rho_{\psi^j}(x^j, A) \geq \rho_{\psi^{j+1}}(x^j, A) \geq \rho_{\psi^{j+1}}(y^j, A) \geq \rho_{\psi^j}(y^j, A)$ holds. Given that the intra-personal heterogeneity within menu A is equal to $1 - \sum_{z \in A} \rho^2_{\psi^j}(z, A)$, the transfer must increase the heterogeneity of menu A. Additivity across menus guarantees that $\mathit{Intra}_\lambda(\psi^{j+1}) \geq \mathit{Intra}_\lambda(\psi^j)$. The recursive application of this argument over the sequence of RUMs concludes the proof.■

Proof of Proposition 7: Consider any menu $A \in \mathcal{A}$ and denote its alternatives as $\{y_k\}_{k=1}^K$ with the property that $u_1(y_1) \geq \cdots \geq u_1(y_K)$ and $u_2(y_1) \geq \cdots \geq u_2(y_K)$. First, notice that the assumption guarantees that $\frac{u_2(y_s)}{u_2(y_t)} \geq \frac{u_1(y_s)}{u_1(y_t)}$ for every $s > t$ and, hence, $\frac{\rho_{\psi_{u_2}}(y_s,A)}{\rho_{\psi_{u_2}}(y_t,A)} = \frac{\frac{u_2(y_s)}{\sum_{k=1}^K u_2(y_k)}}{\frac{u_2(y_t)}{\sum_{k=1}^K u_2(y_k)}} \geq \frac{\frac{u_1(y_s)}{\sum_{k=1}^K u_1(y_k)}}{\frac{u_1(y_t)}{\sum_{k=1}^K u_1(y_k)}} = \frac{\rho_{\psi_{u_1}}(y_s,A)}{\rho_{\psi_{u_1}}(y_t,A)}$. That is, the choice probabilities in menu A are also related by the monotone likelihood ratio property. As a result, we know that there exists $T \leq K$ such that $\rho_{\psi_{u_1}}(y_t, A) \geq \rho_{\psi_{u_2}}(y_t, A)$ if and only if $t \leq T$. Since $\sum_{k=1}^K \rho_{\psi_{u_1}}(y_k, A) = \sum_{k=1}^K \rho_{\psi_{u_2}}(y_k, A) = 1$, the uniform distribution over $\{\rho_{\psi_{u_2}}(y_k, A)\}_{k=1}^K$ second-order stochastically dominates the uniform distribution over $\{\rho_{\psi_{u_1}}(y_k, A)\}_{k=1}^K$. The strict convexity of the quadratic function guarantees that $\frac{\sum_{k=1}^K(\rho_{\psi_{u_1}}(y_k,A))^2}{K} \geq \frac{\sum_{k=1}^K(\rho_{\psi_{u_2}}(y_k,A))^2}{K}$, or equivalently $\sum_{k=1}^K(\rho_{\psi_{u_1}}(y_k, A))^2 \geq \sum_{k=1}^K(\rho_{\psi_{u_2}}(y_k, A))^2$. Conditional on menu $A \in \mathcal{A}$, we can write intra-personal heterogeneity as 1 minus the previous sums of squares and, hence, the heterogeneity within menu A is larger for the Luce RUM defined by v. Additivity of intra-personal heterogeneity across menus concludes the proof. ■

Proof of Corollary 4: From Corollary 3, $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1; \psi_{\mathcal{U}}]) = \alpha \mathsf{CH}_\lambda(\theta) + (1-\alpha)\mathsf{CH}_\lambda([1; \psi_{\mathcal{U}}]) + \alpha(1-\alpha) d_\lambda(\psi_\theta, \psi_{\mathcal{U}})$. From Proposition 4, this is equivalent to $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1; \psi_{\mathcal{U}}]) = \alpha(\beta_\lambda - d_\lambda(\psi_\theta, \psi_{\mathcal{U}})) + (1-\alpha)\beta_\lambda + \alpha(1-\alpha) d_\lambda(\psi_\theta, \psi_{\mathcal{U}}) = \beta_\lambda - \alpha^2 d_\lambda(\psi_\theta, \psi_{\mathcal{U}})$. ■

References

[1] Apesteguia, J. and M.A. Ballester. (2021). “Separating Predicted Randomness from Residual Behavior.” *Journal of the European Economic Association*, 19(2), 1041-1076.

[2] Baldiga, K. A. and J. R. Green. (2013). “Assent-maximizing social choice.” *Social Choice and Welfare*, 40(2), 439-460.

[3] Barseghyan, L., F. Molinari, and M. Thirkettle. (2021). “Discrete choice under risk with limited consideration.” *American Economic Review*, 111(6), 1972-2006.

[4] Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and C.J. Tyson. (2022). “Mixture Choice Data: Revealing Preferences and Cognition.” *Journal of Political Economy*, forthcoming.

[5] de Clippel, G. and K. Rozen. (2022). “Which Performs Best? Comparing Discrete Choice Models.” Mimeo.

[6] Ely, J., A. Frankel and E. Kamenica. (2015). “Suspense and surprise.” *Journal of Political Economy*, 123(1), 215-260.

[7] Esteban, J. M. and D. Ray. (1994). “On the measurement of polarization.” *Econometrica*, 62(4), 819-851.

[8] Frankel, D. M. and O. Volij, O. (2011). “Measuring school segregation.” *Journal of Economic Theory*, 146(1), 1-38.

[9] Gentzkow, M., J. M. Shapiro and M. Taddy. (2019). “Measuring group differences in high-dimensional choices: method and application to congressional speech.” *Econometrica*, 87(4), 1307-1340.

[10] Greenberg, J.H. (1956). “The measurement of linguistic diversity.” *Language*, 32(1), 109-115.

[11] Halevy, Y., D. Persitz, and L. Zrill. (2018). “Parametric recoverability of preferences.” *Journal of Political Economy*, 126(4), 1558-1593.

[12] Kitamura, Y., and J. Stoye. (2018). “Nonparametric analysis of random utility models.” *Econometrica*, 86(6), 1883-1909.

[13] Leonhardt, U. (1997). *Measuring the quantum state of light.* Cambridge University Press.

[14] Lieberson, S. (1969). “Measuring population diversity.” *American Sociological Review*, 34(6), 850-862.

[15] Simpson, E.H. (1949). “Measurement of diversity.” *Nature*, 163(4148), 688-688.

[16] Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge University Press.