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 randomly selected 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

In this paper, we provide a way to measure behavioral heterogeneity, which is, by now, a well-established phenomenon in economics. Ultimately, measuring heterogeneity will allow for a thorough understanding of its causes and implications. For example, measuring heterogeneity is essential for comprehending its underlying determinants, such as demographics, education, or rationality. It can also enhance prediction exercises, as lower heterogeneity is expected to improve predictive accuracy. Additionally, it is a crucial step in developing a representative stochastic-agent model that captures population variability. Lastly, accounting for heterogeneity is vital 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

Date: May, 2023.

*We thank Larbi Alaoui, David Jimenez-Gomez, Damien Mayaux, Andrea Salvanti, Jesse Shapiro and Rafael Suchy for helpful comments. Financial support by FEDER/Ministerio de Ciencia e Innovación (Agencia Estatal de Investigación) through Grant PID2021-125538NB-I00 and through the Severo Ochoa Programme for Centers 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.

vary in their tastes and, therefore, in their economic choices. Second, the behavior of 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.

Given its prevalence in theoretical and applied work, we adopt a random utility framework.[1] To allow for the possibility of both inter- and intra-personal variability, we formalize an individual as a random utility model and a population as a distribution over such individuals. Then, 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, that we refer to by CH. This measure of heterogeneity aligns well with traditional diversity measurement in various fields, as discussed in Section 2, and thus it is a natural starting point.

In Section 4 we discuss four convenient features of CH. First, we prove that CH can be computed even when there is only population aggregate data, a limitation often faced by the analyst. Second, we obtain a matrix representation of CH that emphasizes that it is easily implementable in practice. Third, we establish that the measure can be equivalently derived as a Euclidean distance in the space of choice functions. Finally, by utilizing this Euclidean representation, we demonstrate that CH enables a convenient differentiation between inter- and intra-personal components, which can be valuable in panel data analysis.

In Section 5 we consider properties of a heterogeneity measure with the ultimate goal of providing axiomatic foundations for CH. The first property is a reduction principle, establishing that heterogeneity can be computed using aggregate choice data. The second property is a decomposition principle, stating that heterogeneity is computed as a weighted sum of the heterogeneity of populations consisting of two deterministic individuals. Finally, the third property is a monotonicity principle by which an increase in choice divergence augments heterogeneity. Theorem 1 provides a characterization of CH based on these three properties.

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

Having proposed and studied our choice-based measure of behavioral heterogeneity, in Section 6 we further elaborate on the comparative statics of the two components of heterogeneity. We start with intra-personal heterogeneity to show that when considering individuals with a central preference relation, moving mass away from preferences that are closer to the central preference relation increases intra-personal heterogeneity. In terms of inter-personal heterogeneity, we show that mixing a given population with another population with larger heterogeneity increases overall heterogeneity due to the added inter-personal variability.

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 paper contributes by proposing an overall measure of heterogeneity that applies to settings where there are two layers, inter- and intra-personal, of heterogeneity. In addition, we are concerned with choice behavior, which involves a number of overlapping situations (i.e., choices from not just one, but different menus), and we provide axiomatic foundations.

Economics uses a number of alternative approaches for measuring inter-personal preference 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) provides a choice-based analysis of consensus, and Gentzkow, Shapiro and Taddy (2019) studies partisanship based on the predictability of party speeches. We contribute to this literature by providing a measure of both intra- and inter-personal behavioral heterogeneity within a unique choice framework.

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 contribute to this literature by offering a measure of heterogeneity based on first principles.

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. An individual ψ is formalized as a random utility model; that is, ψ is a probability distribution on $\mathcal{P}$, 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, individual choices are stochastic. Denoting by $m(A, P)$ the maximal alternative in menu A according to preference P, and by $\mathbb{I}_{[.]}$ the indicator function which takes the value 1 when the statement in brackets is true and 0 otherwise, the probability that individual 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 possible individuals and by Ψ^D the set of all individuals that are deterministic, i.e., that assign mass 1 to a single preference. For the latter class, we denote by ψ_P the deterministic individual associated to preference P. In addition, we denote by $\psi_{\mathcal{U}}$ the (uniform) individual that assigns equal mass to all preferences.

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

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

with θ_i describing the mass of individual ψ_i in the population, and $\sum_i \theta_i = 1$. We denote by Θ the set of all possible populations and by Θ^D the set of all deterministic populations, i.e., those with the form $[\theta_1, \theta_2, \ldots, \theta_m; \psi_{P_1}, \psi_{P_2}, \ldots, \psi_{P_m}]$, which assign mass only to deterministic individuals. In words, a deterministic population represents the case of a population in which all individuals are deterministic but possibly heterogeneous. Alternatively, denote by Θ^{hom} the set of all populations that are homogeneous, i.e., taking the form $[1; \psi]$. That is, a homogeneous population represents

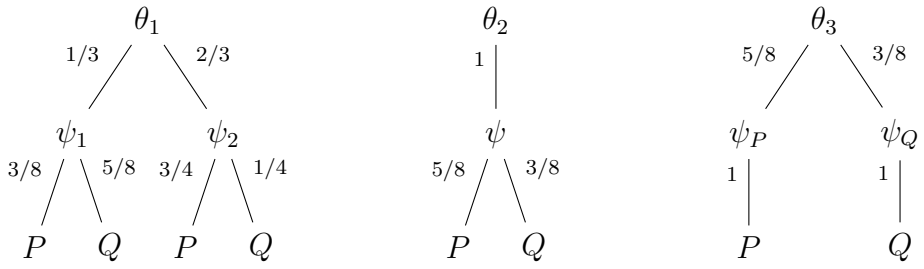
[2] Given the relevance of the Luce and mixed-logit models in applications, we use them to illustrate some of our results.

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

the 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 individual 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 individuals, given 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 are non-deterministic, placing 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 individuals, ψ_P and ψ_Q, with masses $\frac{5}{8}$ and $\frac{3}{8}$ respectively. □

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 $\mathsf{CH}_\lambda(\theta^1) = \frac{1}{3}[\frac{1}{3}(\frac{3}{8}\frac{5}{8} + \frac{5}{8}\frac{3}{8}) + \frac{2}{3}(\frac{3}{8}\frac{1}{4} + \frac{5}{8}\frac{3}{4})] + \frac{2}{3}[\frac{1}{3}(\frac{3}{4}\frac{5}{8} + \frac{1}{4}\frac{3}{8}) + \frac{2}{3}(\frac{3}{4}\frac{1}{4} + \frac{1}{4}\frac{3}{4})] = \frac{15}{32}$. □

Notice that CH_λ establishes a complete and transitive ranking of behavioral heterogeneity on the space of all populations.[5] We now discuss four results on the structure of CH_λ, that may be attractive in certain settings. The first emphasizes the fact that CH_λ can be computed even when there is only population aggregate data. The second uses a matrix representation that shows the computational convenience of the measure. The third relates CH_λ to a Euclidean distance, connecting the measure with standard practices in econometric estimations. Finally, we show that CH_λ allows for a convenient distinction between inter- and intra-personal components, that may be of use in the presence of panel data.

4.1. **Aggregate data.** $\mathsf{CH}_\lambda(\theta)$ is formally defined using panel data, with information on the choices ρ_{ψ_i} of every individual in population θ. However, it is often the case that choice data is only available in aggregate terms, i.e., in the weighted average form given by $\sum_i \theta_i \rho_{\psi_i}$. The question arises on whether computing the behavioral heterogeneity using aggregate data gives the same overall choice heterogeneity than if one would had panel data. Below we show that the answer to this question is positive. To formalize this result, notice that Ψ is convex and thus, aggregate data can be seen as produced by a homogeneous population where every individual behaves like, what we call, the representative agent $\psi_\theta = \sum_i \theta_i \psi_i$. We then have the following result.[6]

Proposition 1. $\mathsf{CH}_\lambda(\theta) = \mathsf{CH}_\lambda([1;\psi_\theta])$.

Example 1 (continued). The representative agent 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 $\mathsf{CH}_\lambda(\theta^2) = \frac{5}{8}\frac{3}{8} + \frac{3}{8}\frac{5}{8} = \frac{15}{32} = \mathsf{CH}_\lambda(\theta^1)$. □

4.2. **A matrix representation of CH.** We now show that we can use the representative agent of the population, together with an account of the heterogeneity of simple

[5]Therefore, the analysis of choice-based heterogeneity could be equivalently described in terms of a complete and transitive binary relation over the space of populations.

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

populations, in order to provide a convenient matrix representation of our measure of heterogeneity.

We refer to a population composed exclusively of two equally-weighted deterministic individuals, $[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$, as a couple. Now, compile twice the heterogeneity value of each possible couple in a $|\mathcal{P}| \times |\mathcal{P}|$-matrix that we denote by $\mathcal{C}_\lambda$. Note that this is a symmetric matrix with zeros in the diagonal and the entry for a given couple equal to the sum of the λ-weights of the menus where its two individuals differ in their choices. It is important to stress that this matrix is independent of the specific distribution over the individuals, and hence independent of the population, since it is characterized by the choice disagreements between preferences, weighted by measure λ. Therefore, given $\mathcal{P}$, the matrix does not need to be recalculated for the analysis of different populations, or for behavioral variations within a population, which is a computationally convenient property in practice.

Example 2.[7] Let $X = \{x, y, z\}$ and the distribution over menus $\bar{\lambda}$ 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

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

□

Proposition 2 shows that the choice heterogeneity of any population can be seen as an inner product involving its representative agent and matrix $\mathcal{C}_\lambda$.[8]

Proposition 2. $\mathsf{CH}_\lambda(\theta) = \psi_\theta \mathcal{C}_\lambda \psi_\theta^\top$.

Example 2 (continued). We consider here the case of mixed-logit populations $\theta = [\theta_1, \theta_2, \ldots, \theta_m; u_1, u_2, \ldots, u_m]$, where each individual corresponds to a Luce model.[9]

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

[8] This is due to the fact that $\mathcal{C}_\lambda$ is a symmetric positive semi-definite matrix, admitting a Cholesky factorization.

[9] A Luce model 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 $\sum_{x \in X} u(x) = 1$ and hence, $u(x)$ can be understood as the probability of choosing x in X and, for every menu A, individual choice probabilities are simply conditional probabilities. Luce models

Given preference P, described by $x_1 P x_2 P \dots x_{N-1} P x_N$, the probability assigned by the representative agent of the mixed-logit population is $\psi_\theta(P) = \sum_i \theta_i \prod_{j=1}^{N} \frac{u_i(x_j)}{\sum_{k=j}^{N} u_i(x_k)}$. Following Example 2, consider, e.g., $\theta = [\frac{4}{11}, \frac{7}{11}; \psi_1, \psi_2]$, with $u_1 = (1/2, 1/3, 1/6)$, and $u_2 = (4/9, 3/9, 2/9)$. The representative agent is $\psi_\theta = \frac{1}{495}(144, 86, 115, 50, 58, 42)$ and hence, $\mathsf{CH}_{\bar{\lambda}}(\theta) = \psi_\theta \mathcal{C}_{\bar{\lambda}} \psi_\theta^\top = .5$. □

4.3. **A 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 agent and the stochastic choice function providing maximal heterogeneity, that is the one given by uniformly random behavior.[10]

Formally, given any two individuals 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 3. $\mathsf{CH}_\lambda(\theta) = \beta_\lambda - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}}) = \max_{\psi \in \Psi} d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}}) - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}})$
$= d_\lambda(\rho_{\psi_P}, \rho_{\psi_\mathcal{U}}) - d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}})$ *for every* $P \in \mathcal{P}$.

Proposition 3 first shows that the choice heterogeneity of a population is inversely related to the distance between the stochastic choice function of its representative agent and uniform choices. Moreover, the second equality in Proposition 3 shows that the constant β_λ is in fact the maximum possible distance between an individual and uniform choices, and the third equality establishes that this corresponds to the distance between any deterministic individual and uniform choices.

Example 1 (continued). Since in Example 1 there is only one binary menu, $\beta_\lambda = \frac{1}{2}$. Now, using our convention to represent individuals 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 $\mathsf{CH}_\lambda(\theta^1) = \frac{15}{32} = \frac{1}{2} - [(\frac{5}{8} - \frac{1}{2})^2 + (\frac{3}{8} - \frac{1}{2})^2]$. □

admit different RUM representations but, since all of them generate the same stochastic choices, this is inconsequential for our analysis and we will simply write u_i instead of ψ_i whenever needed.

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

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

Proposition 4. $\mathsf{CH}_\lambda(\theta) = \sum_i \theta_i[\beta_\lambda - d_\lambda(\rho_{\psi_i}, \rho_{\psi_\mathcal{U}})] + \sum_i \theta_i \sum_{i<j} \theta_j \; d_\lambda(\rho_{\psi_i}, \rho_{\psi_j})$.

Proposition 4 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. This term, then, aggregates only intra-personal variability across the individuals 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. Accordingly, this second term measures only inter-personal variability between the members of the population.

Example 1 (continued). 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 $\mathsf{CH}_\lambda(\theta_1) = \frac{1}{3}(\frac{1}{2} - \frac{1}{32}) + \frac{2}{3}(\frac{1}{2} - \frac{1}{8}) + \frac{1}{3}\frac{2}{3}\frac{9}{32} = \frac{15}{32}$. □

In Section 6 we return to this decomposition and study formally each of the two components of CH_λ.

5. A characterization of CH

We now discuss three plausible properties for a measure of behavioral heterogeneity and show they are necessary and sufficient for CH_λ. 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.

The first axiom, Reduction, is the formalization of the ideas discussed in Section 4.1 regarding aggregate data.

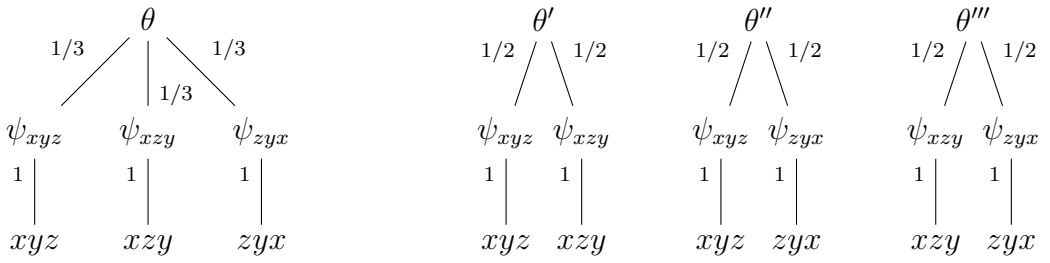
Reduction. $\mathsf{H}(\theta) = \mathsf{H}([1; \psi_\theta])$.

For our next axiom, consider a deterministic population $\theta \in \Theta^D$. We study 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 individuals, with weights in proportion to their masses in the original population, i.e., sub-populations of the form $[\frac{\theta_i}{\theta_i+\theta_j}, \frac{\theta_j}{\theta_i+\theta_j}; \psi_{P_i}, \psi_{P_j}]$. 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, $(\theta_i + \theta_j)^2$. This leads us to the following property:

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

Example 3. 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.

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

Finally, we discuss a monotonicity property involving only couples. Let us start by defining collections of couples $C = \{[\frac{1}{2}, \frac{1}{2}; \psi_{P^n}, \psi_{Q^n}]\}_{n=1}^N$. Now, consider two equally-sized collections of couples C and C', that is $N = N'$, and suppose that whatever the menu at hand, we unequivocally observe a larger number of choice-disagreements in C than in C'. In such a case, it is natural to conclude that the average heterogeneity of C 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. Then:

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 $\overline{\mathsf{H}}(C) \geq \overline{\mathsf{H}}(C')$.

Example 3 (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 collections C and C' must be equal, and lower than the average heterogeneity of collection C''. Indeed, our computation above showed that the average heterogeneity of C is $\frac{1}{3}$. 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 $\mathsf{CH}_\lambda([\frac{1}{3}, \frac{1}{3}, \frac{1}{3}; \psi_{xyz}, \psi_{xzy}, \psi_{zyx}]) = \mathsf{CH}_\lambda([\frac{1}{3}, \frac{1}{3}, \frac{1}{3}; \psi_{xyz}, \psi_{zxy}, \psi_{zyx}]) \leq \mathsf{CH}_\lambda([\frac{1}{3}, \frac{1}{3}, \frac{1}{3}; \psi_{xyz}, \psi_{yxz}, \psi_{zyx}])$. □

We can now establish the following characterization result.

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 agent $[1, \psi_\theta]$. Thus, we consider the deterministic population θ^d that assigns the same probability to every preference as the representative agent of θ.[11] Hence, since θ and θ^d have the same representative agent, 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, $[\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$. The function is actually the logistic map which yields

[11] Note that in Example 1, this deterministic population corresponds to population θ^3.

$\mathsf{H}([1-\gamma,\gamma;\psi_P,\psi_Q]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2},\frac{1}{2};\psi_P,\psi_Q])$. 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.[12]

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 over 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.

6. Comparative statics: Intra- and inter-personal heterogeneity

We now build on the decomposition obtained in Section 4.4 to establish further results with respect to intra- and inter-personal heterogeneity.

6.1. **Intra-personal heterogeneity.**

Given an individual ψ, it is natural to assess its intra-personal heterogeneity.[13] Proposition 4 shows that we can do so by way of the λ-Euclidean distance between individual behavior and uniform choices, $d_\lambda(\rho_\psi, \rho_{\psi_\mathcal{U}})$.

We now investigate further the structure of intra-personal heterogeneity. For this, we use a particular class of individual behaviors, namely, those for which there is a central preference and, in every menu, better alternatives are chosen with larger probability. Formally, for a given $P \in \mathcal{P}$, we say that ψ is P-central if xPy and $\{x,y\} \subseteq A$ implies $\rho_\psi(x,A) \geq \rho_\psi(y,A)$. The notion of P-centrality is related to the well-known notion of weak stochastic transitivity. Any P-central individual 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 individuals is the Luce model, as well as many of its generalizations.

Given two P-central individuals, 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)

[12] This provides the type of decomposition described in Proposition 4.

[13] We are agnostic as for the interpretation of intra-personal variability. For discussions on the possible connection between rationality and intra-personal heterogeneity see Apesteguia and Ballester (2015, 2021) and Ok and Tserenjigmid (2023).

$\psi_2 = \psi_1 - \epsilon\psi_{Q_1} + \epsilon\psi_{Q_2}$ and (ii) Q_2 is farther away from P than Q_1 is, i.e., xPy and xQ_2y imply xQ_1y. That is, the second individual is obtained from the first by shifting mass from preference Q_1 to preference Q_2, which happens to be farther away from the central preference P. Proposition 5 shows that, in accordance with intuition, this type of shift increases intra-personal heterogeneity. Indeed, the result is also true when sequential changes 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.[14]

Proposition 5. *If ψ_2 is a sequential decentralization of ψ_1, $d_\lambda(\rho_{\psi_1}, \rho_{\psi_\mathcal{U}}) \geq d_\lambda(\rho_{\psi_2}, \rho_{\psi_\mathcal{U}})$.*

Proposition 5 establishes some intuitive comparative statics on intra-personal heterogeneity for P-central individuals. 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.[15]

Proposition 6. *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 $i < j$, $d_\lambda(\rho_{\psi_{u_1}}, \rho_{\psi_\mathcal{U}}) \geq d_\lambda(\rho_{\psi_{u_2}}, \rho_{\psi_\mathcal{U}})$.*

Proposition 6 considers two Luce individuals with the same central preference. By the monotone likelihood ratio, u_2 places more mass on worse alternatives, and hence Proposition 6 establishes that it must have a larger amount of intra-personal heterogeneity.

Example 2 (continued). Since the monotone likelihood ratio holds for u_1 and u_2, Proposition 6 implies that $d_\lambda(\rho_{\psi_{u_1}}, \rho_{\psi_\mathcal{U}}) \geq d_\lambda(\rho_{\psi_{u_2}}, \rho_{\psi_\mathcal{U}})$. Since $\psi_{u_1} = \frac{1}{60}(20, 10, 15, 5, 6, 4)$ and $\psi_{u_2} = \frac{1}{315}(84, 56, 70, 35, 40, 30)$, it follows immediately that $d_\lambda(\rho_{\psi_{u_1}}, \rho_{\psi_\mathcal{U}}) = .1$ and $d_\lambda(\rho_{\psi_{u_2}}, \rho_{\psi_\mathcal{U}}) = .07$. Consider now the representative agent ψ_θ. Since this is not a Luce individual, Proposition 6 cannot be applied. However, ψ_θ happens to be a P-central individual, 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 5 implies that $d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}}) \in [d_\lambda(\rho_{\psi_{u_2}}, \rho_{\psi_\mathcal{U}}), d_\lambda(\rho_{\psi_{u_1}}, \rho_{\psi_\mathcal{U}})]$. Notice that $d_\lambda(\rho_{\psi_\theta}, \rho_{\psi_\mathcal{U}}) = .08$, consistent with the claim. □

[14] 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.

[15] The required notation is given in Example 2.

6.2. **Inter-personal heterogeneity.** Proposition 4 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 case because the reasoning in Proposition 4 can be extended to combinations of any two populations θ and θ'.[16]

Corollary 1. *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 1 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 agents of the sup-populations. 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 1 confirms this by showing that the additional heterogeneity can be obtained simply by inspecting the distance between the representative agents.

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,

Proposition 7. *For every* $\alpha \in [0,1]$, $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1;\psi_{\mathcal{U}}]) = \beta_\lambda - \alpha^2 d_\lambda(\psi_\theta, \psi_{\mathcal{U}})$.

[16] 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 $1-\alpha$.

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 1 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 Proposition 7, is $\frac{1}{2} - \alpha^2\frac{1}{32}$, a value that increases with the trembling weight $1-\alpha$. □

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 4 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 convex and, for any given menu, it is possible to construct a pair of deterministic choice functions that differ only over the given menu. Hence, our characterization result can be adapted to this setting.

Our modeling of individual behavior implicitly assumes that individual choices are independent. One may be interested in introducing the possibility of correlated choices. This can be incorporated into our framework by considering state-dependent preferences. That is, there is a common set of states across individuals and a common probability distribution over them, and each individual is described by a mapping from states to preferences. In this setting, choice heterogeneity could be measured by the probability that the choices of two sampled individuals differ over a sampled state within a sampled menu.

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., Agranov and Ortoleva (2017), Halevy, Persitz, and Zrill (2018), Kitamura and Stoye (2018), Natenzon (2019), Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), Fudenberg, Newey, Strack, and Strzalecki (2020), Aguiar and Kashaev (2021), Alós-Ferrer, Fehr, and Netzer (2021), Apesteguia and Ballester (2021), Barseghyan, Molinari, and Thirkettle (2021), Caplin and Martin (2021), Dardanoni, Manzini, Mariotti, Petri, and Tyson (2022), Dean, Ravindran, and Stoye (2022), de Clippel and Rozen (2022), and Jagelka (2023)). Once the individual stochastic models are specified, the application of our measure is direct, as discussed in the main text (see, in particular, Section 6).

Appendix A. Proofs

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

$$\begin{aligned}
\mathsf{CH}_\lambda(\theta) &= \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_a \rho_{\psi_i}(a,A)(1-\rho_{\psi_j}(a,A)) \\
&= \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_P \psi_i(P) \sum_Q \psi_j(Q) \cdot \mathbb{I}_{[m(A,P)\neq m(A,Q)]} \\
&= \sum_A \lambda(A) \sum_i \sum_P \theta_i \psi_i(P) \sum_j \sum_Q \theta_j \psi_j(Q) \cdot \mathbb{I}_{[m(A,P)\neq m(A,Q)]} \\
&= \sum_A \lambda(A) \sum_P \psi_\theta(P) \sum_Q \psi_\theta(Q) \cdot \mathbb{I}_{[m(A,P)\neq m(A,Q)]} \\
&= \sum_A \lambda(A) \sum_a \rho_{\psi_\theta}(a,A)(1-\rho_{\psi_\theta}(a,A)) = \mathsf{CH}_\lambda([1;\psi_\theta]).
\end{aligned}$$

■

Proof of Proposition 2: The proof follows from the proof of Theorem 1. ■

Proof of Proposition 3: We start by proving a series of useful claims. The first is that, conditional on having sampled the ordered pair of individuals (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 individuals (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

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

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

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 individuals weighted by their corresponding sampling probabilities. Hence, we proceed by aggregating the expression given above. Since every individual ψ_i appears as the first individual in the sampling with probability θ_i and again, as the second individual 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 individuals 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 individual 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 individuals 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:

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

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

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 individual 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 4: The proof follows directly from the second and third claims in the proof of Proposition 3. ■

Proof of Theorem 1: The necessity of Reduction is shown in Proposition 1. For Decomposition, 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 individual makes two different choices is zero, and hence the heterogeneity of θ can be written as

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

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

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

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

For Monotonicity, note that the average heterogeneity of $C = \{[\frac{1}{2}, \frac{1}{2}; \psi_{P^n}, \psi_{Q^n}]\}_{n=1}^N$ 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 necessity of Monotonicity follows.

Finally, 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 part. 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

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

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 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 $\{P_{B'}'^A, Q_{B'}'^A\}_{B':\{a',b'\}\subseteq B'\subseteq A}$. 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 $[\frac{1}{2}, \frac{1}{2}; \psi_{P_{B'}'^A}, \psi_{Q_{B'}'^A}]$ where $\{a', b'\} \subseteq B' \subseteq A$ satisfies $(-1)^{|A|-|B'|} = 1$ and, finally (iv) $C_2'^A$ is formed by all couples $[\frac{1}{2}, \frac{1}{2}; \psi_{P_{B'}'^A}, \psi_{Q_{B'}'^A}]$ 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

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

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. Given that $\mathsf{H}([\frac{1}{2},\frac{1}{2};\psi_R,\psi_R]) = 0$, we can 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 $\mathsf{H}([1-\gamma, \gamma; \psi_P, \psi_Q]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_P, \psi_Q]$.

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 agent of θ). From Claim 5, $\mathsf{H}(\theta^d) = \sum_{i<j} 4\,\theta^d_i\theta^d_j\mathsf{H}([\frac{1}{2}, \frac{1}{2}; \psi_{P_i}, \psi_{P_j}])$. Using

Claim 2, we have $\mathsf{H}(\theta^d) = \sum_{i<j} 4\,\theta_i^d\theta_j^d \sum_{A:m(A,P_i)\neq m(A,P_j)} \tau(A)$. 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 agent 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 5: Suppose that ψ_2 is a sequential decentralization of ψ_1. By definition, there exists a sequence $\{\psi^j\}_{j=1}^J$ of individuals such that $\psi^1 = \psi_1$ and $\psi^J = \psi_2$, and ψ^j is a decentralization of ψ^{j-1} for $j = 2,\ldots,J$, 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 $x^jPy^j, x^jQ_1^jy^j$ and $y^jQ_2^jx^j$.

First, consider any menu A that does not 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 heterogeneity of population $[1;\psi^j]$ within menu A is equal to $1-\sum_{z\in A}\rho_{\psi^j}^2(z,A)$, the transfer must increase the heterogeneity of menu A. Additivity across menus guarantees that $\mathsf{CH}_\lambda([1;\psi^{j+1}]) \geq \mathsf{CH}_\lambda([1;\psi^j])$. The recursive application of this argument over the sequence of individuals together with Proposition 3 concludes the proof. ■

Proof of Proposition 6: 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 defined by v. Additivity of intra-personal heterogeneity across menus concludes the proof. ■

Proof of Proposition 7: From Corollary 1, $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1;\psi_{\mathcal{U}}]) = \alpha\mathsf{CH}_\lambda(\theta) + (1-\alpha)\mathsf{CH}_\lambda([1;\psi_{\mathcal{U}}]) + \alpha(1-\alpha)d_\lambda(\psi_\theta, \psi_{\mathcal{U}})$. From Proposition 3, 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] Agranov, M. and P. Ortoleva (2017). "Stochastic choice and preferences for randomization." *Journal of Political Economy*, 125(1), 40-68.

[2] Aguiar, V. H. and N. Kashaev (2021). "Stochastic revealed preferences with measurement error." *Review of Economic Studies*, 88(4), 2042-2093.

[3] Alós-Ferrer, C., E. Fehr, and N. Netzer (2021). "Time will tell: Recovering preferences when choices are noisy." *Journal of Political Economy*, 129(6), 1828-1877.

[4] Apesteguia, J. and M. A. Ballester (2015). "A measure of rationality and welfare." *Journal of Political Economy*, 123(6), 1278-1310.

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

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

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

[8] Caplin, A. and D. Martin (2021). "Comparison of decisions under unknown experiments." *Journal of Political Economy*, 129(11), 3185-3205.

[9] 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.

[10] Dean, M., D. Ravindran, and J. Stoye (2022). "A Better Test of Choice Overload." Mimeo.

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

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

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

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

[15] Fudenberg, D., Newey, W., Strack, P., and T. Strzalecki (2020). "Testing the drift-diffusion model." *Proceedings of the National Academy of Sciences*, 117(52), 33141-33148.

[16] 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.

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

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

[19] Jagelka, T. (2023). "Are economists' preferences psychologists' personality traits? A structural approach." *Journal of Political Economy*, forthcoming.

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

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

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

[23] Natenzon, P. (2019). "Random choice and learning." *Journal of Political Economy*, 127(1), 419-457.

[24] Ok, E. A., and G. Tserenjigmid (2023). "Measuring Stochastic Rationality." Mimeo.

[25] Simpson, E.H. (1949). "Measurement of diversity." *Nature*, 163(4148), 688-688.

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