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 allows for a more thorough understanding of its causes and implications. For example, it is essential for identifying 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.

We start by formalizing an individual, or type, by means of a stochastic choice function, and a population as a distribution over such individuals. We do so because the behavioral heterogeneity of a population may be the result of two different phenomena.

Date: February, 2025.

*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`.

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 any given individual is also subject to variation. We then 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 discuss in Section 2, and thus it is a natural starting point. Moreover, as we discussed in Section 4, CH can be instrumental in the study of a variety of economic applications. In particular, we suggest two general problems involving externalities, with the first one emphasizing inter-temporal factors, and the second one emphasizing spatial factors.

In Section 5 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 establish that the measure can be equivalently derived as a Euclidean distance in the space of choice functions. Third, 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; 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. Finally, we use the previous decomposition to analyze the heterogeneity of combinations of any two populations.

In Section 6 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.

The most influential model in the stochastic choice literature entails the random utility model (RUM). Accordingly, in Section 7, we study the implementation of our measure to the case of populations where types are described as RUMs. We show that the RUM framework enhances the tractability of computing heterogeneity, facilitates simple comparative statics involving intra-personal heterogeneity, and allows us to adapt the axiomatic treatment from Section 6 to the RUM case.

Section 8 concludes with comments on various possible extensions of the measure and its empirical implementation.

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, Gentzkow, Shapiro and Taddy (2019) studies partisanship based on the predictability of party speeches, and Bertrand and Kamenica (2023) analyzes temporal trades in cultural distances between groups. 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 stochastic choice 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 Luce behaviors is entertained (see Train, 2009).[1] Notably, Dardanoni, Manzini, Mariotti, Petri, and Tyson (2023) study the identification of the type-conditional stochastic choice functions and the type probabilities producing a given aggregate choice data. We contribute to this literature by offering a measure of heterogeneity for this type of populations based on first principles.

3. The measure

Consider a finite set of alternatives X, and let $\mathcal{A}$ be the collection of all subsets of X containing at least two alternatives, which we refer to as menus. Our approach defines a family of heterogeneity measures that depend on a single parameter λ, chosen by the analyst. The parameter λ represents a probability distribution over menus, with $\lambda(A)$ indicating the probability that menu A is sampled. This assumption is irrelevant in applications where behavior is observed for a single menu but can be useful otherwise; when behavior is observed across multiple menus, the analyst may assign greater weight to behavioral divergence in certain menus. This may be appropriate when some menus are encountered more frequently by the population or when the specifics of the application suggest that certain menus are more relevant than others. Importantly, all our results apply to every heterogeneity measure in this family.

An individual, or type, i is described as a stochastic choice function ρ_i such that $\rho_i(a, A) \in [0, 1]$ denotes the probability of choosing a from A and, as usual, $\rho_i(a, A) = 0$ whenever $a \notin A$ and $\sum_{a \in A} \rho_i(a, A) = 1$.[2] A population is a probability distribution over types that assigns strictly positive mass to only a finite number of them, i.e., an object with the form

$$\theta = [\theta_1, \theta_2, \ldots, \theta_m; \rho_1, \rho_2, \ldots, \rho_m],$$

with θ_i describing the mass of type ρ_i in the population, and $\sum_i \theta_i = 1$.

We measure heterogeneity as the probability that, over a sampled menu, the sampled choices of two sampled types differ. Formally, given λ, the choice heterogeneity of population θ is:

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

[1]Given the relevance of the Luce and mixed-logit models in applications, we use them in Section 7 to illustrate some of our results.

[2]In occasions we write $\rho_i(A)$ to refer to the vector of choice probabilities in A given by ρ_i.

Note that different values of λ generate different rankings of heterogeneity. We illustrate the setting and the measure in the following example.

Example 1. Consider the binary set $X = \{x, y\}$. Hence, it must be that $\lambda(\{x, y\}) = 1$, and all that matters in order to describe a type is its probability of choosing x from $\{x, y\}$. Consider the two-types population $\theta = [\frac{1}{3}, \frac{2}{3}; \rho_1, \rho_2]$, where $\rho_1(x, \{x, y\}) = \frac{3}{8}$ and $\rho_2(x, \{x, y\}) = \frac{3}{4}$. The measure of choice heterogeneity of θ is $\mathsf{CH}_\lambda(\theta) = \frac{1}{3}[\frac{1}{3}(\frac{3}{8}\frac{5}{8} + \frac{5}{8}\frac{3}{8}) + \frac{2}{3}(\frac{3}{8}\frac{1}{4} + \frac{5}{8}\frac{3}{4})] + \frac{2}{3}[\frac{1}{3}(\frac{3}{4}\frac{5}{8} + \frac{1}{4}\frac{3}{8}) + \frac{2}{3}(\frac{3}{4}\frac{1}{4} + \frac{1}{4}\frac{3}{4})] = \frac{15}{32}$. $\square$

4. Applications

Our measure provides a tool for the empirical quantification of behavioral heterogeneity of a population which may, in turn, serve as a valuable instrument for understanding its determinants, welfare implications, and related aspects. Beyond this interest, in this section, we show that the measure is critical for understanding a variety of applied economic questions. To illustrate this, we begin by presenting two simple economic settings in which the objective functions coincide with our measure of heterogeneity. Both settings involve the presence of externalities, with the first one emphasizing inter-temporal factors (inter-temporal efficiency, reputation and learning), and the second one emphasizing spatial factors (spatial efficiency, fairness and conflict). We outline a number of economic examples for each case. For clarity, we have limited our discussion to situations where heterogeneity has negative effects, although many other examples involving the positive effects of heterogeneity could also be considered.

4.1. **Sequential Decision-Making.** Consider a population of agents and tasks. At any moment in time, a task A arrives with probability $\lambda(A)$, and agent i responds to this task with probability θ_i. Responses are driven by individual interests, but social externalities exist. Specifically, a unit of social utility is generated only when a decision exactly matches the previous decision for the same task. Consequently, the social value generated by the population θ is captured by $1 - \mathsf{CH}_\lambda(\theta)$. This setting applies to a variety of economic problems. We now present a few examples, organized into three categories.

Inter-temporal efficiency. Stable government regulations minimizes adjustment costs such as administrative ones. In manufacturing processes, consistency in production methods facilitates quality control. Meanwhile, stable working conditions reduce hiring and training costs.

Reputation. Legal rulings that follow consistent principles help establish predictability and fairness in the system, which, in turn, foster social order. In healthcare provision, the use of protocols can increase patients' trust and hence promote adherence of patients to treatments. Likewise, systematic recommendation and support in customer services lead to higher customer satisfaction.

Learning. In academic and parental education, stable teaching methods reinforce knowledge acquisition and hence skill development over time. Also, marketing strategies that emphasize a coherent brand message enhance brand recognition.

4.2. **Spatial Decision-Making.** Consider a population of agents with proportions given by θ_i. Agents are randomly linked according to these proportions. Each agent encounters task A with probability $\lambda(A)$. While responses are driven by individual interests, spatial externalities are present. A unit of social utility is generated only when two neighboring agents, facing the same task, make identical decisions. The expected social utility of the population θ depends on the externalities that are given by $1 - \mathsf{CH}_\lambda(\theta)$. Below we outline some economic examples that fall within this setting, again organized in three categories.

Spatial efficiency. In economic policy-making, regulatory consistency across regions or jurisdictions reduces the bureaucratic burden of trade and improves the efficiency of matching firms with workers. In educational systems, unified regional curricula and standardized testing allow for smoother student transitions and easier nationwide comparisons of learning outcomes, enhancing the efficiency of matching students with universities. Finally, unified legal systems lower the costs of litigation.

Fairness. Examples of critical disparities fostering inequalities and feelings of unfair treatment often involve: different working conditions among sub-units of an organization, different approaches to housing allocation and policing among different municipalities or different taxation and public good provision systems across different regions within a state.

Conflict. Within a firm, differences in work schedules, communication styles, and personal space can create tensions. In residential areas, conflicting lifestyles—such as noise levels at night—often spark disputes. Also, differences in religious observances, cultural practices, social norms or linguistic conventions often amplify misunderstandings and conflict.

5. Structure

We now discuss four results on the structure of CH_λ, each highlighting a different aspect. The first emphasizes the fact that CH_λ can be computed even when there is only aggregate data. The second relates CH_λ to a Euclidean distance, connecting the measure with standard practices in econometric estimations. In the third 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. Finally, the fourth uses the former distinction to facilitate the evaluation of the heterogeneity resulting from the combination of populations.

5.1. **Aggregate data.** CH_λ is formally defined using panel data θ, with information on the proportion θ_i and choices ρ_i of every type i in the population. However, it is often the case that choice data is available only in aggregate terms, raising the question of whether behavioral heterogeneity can be obtained from aggregate data alone. Below we show that the answer to this question is positive. To formalize this result, notice that the space of stochastic choice functions is convex and thus, aggregate data can be seen as produced by a homogeneous population where every type behaves like, what we call, the representative agent $\rho_\theta = \sum_i \theta_i \rho_i$. We then have the following result.[3]

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

Example 1 (continued). The representative agent of population θ is $\rho_\theta(x, \{x, y\}) = \frac{1}{3}\frac{3}{8} + \frac{2}{3}\frac{3}{4} = \frac{5}{8}$. Notice that the direct computation of heterogeneity using the representative agent gives $\mathsf{CH}_\lambda([1; \rho_\theta]) = \frac{5}{8}\frac{3}{8} + \frac{3}{8}\frac{5}{8} = \frac{15}{32} = \mathsf{CH}_\lambda(\theta)$. □

5.2. **A Euclidean representation.** 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.[4]

Formally, given the behavior of any two types ρ and ρ', define the λ-Euclidean distance between their associated stochastic choice functions by

$$\|\rho, \rho'\|_\lambda = \sum_A \lambda(A) \sum_a [\rho(a, A) - \rho'(a, A)]^2.$$

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

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

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. Denote by u the stochastic choice function that, in every menu, assigns the same choice probability to all alternatives, and by d any degenerate stochastic choice function.

Proposition 2. $\mathsf{CH}_\lambda(\theta) = \beta_\lambda - \|\rho_\theta, u\|_\lambda = \max_\rho \|\rho, u\|_\lambda - \|\rho_\theta, u\|_\lambda = \|d, u\|_\lambda - \|\rho_\theta, u\|_\lambda$.

Proposition 2 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 2 shows that the constant β_λ is in fact the maximum possible distance between a type and uniform choices, and the third equality establishes that this corresponds to the distance between any deterministic behavior and uniform choices.

Example 1 (continued). In our streamlined example with just one binary menu, $\beta_\lambda = \frac{1}{2}$ and $u(x, \{x, y\}) = \frac{1}{2}$. Recall that $\rho_\theta(x, \{x, y\}) = \frac{5}{8}$ and hence we can use (the first expression of) Proposition 2 to compute the behavioral heterogeneity of θ, $\mathsf{CH}_\lambda(\theta) = \frac{1}{2} - [(\frac{5}{8} - \frac{1}{2})^2 + (\frac{3}{8} - \frac{1}{2})^2] = \frac{15}{32}$. □

5.3. **A decomposition 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 an intra- and an inter-personal component.

Proposition 3. $\mathsf{CH}_\lambda(\theta) = \sum_i \theta_i[\beta_\lambda - \|\rho_i, u\|_\lambda] + \sum_i \theta_i \sum_{i<j} \theta_j \ \|\rho_i, \rho_j\|_\lambda$.

Proposition 3 shows that choice heterogeneity can be decomposed as the aggregation of two different terms. The first of these terms, $\sum_i \theta_i[\beta_\lambda - \|\rho_i, u\|_\lambda]$, evaluates how close each of the types 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 types in the population. The second term, $\sum_i \theta_i \sum_{i<j} \theta_j \ \|\rho_i, \rho_j\|_\lambda$, evaluates the distance between every pair of different types 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 $\|\rho_1, u\|_\lambda = (\frac{3}{8} - \frac{1}{2})^2 + (\frac{5}{8} - \frac{1}{2})^2 = \frac{1}{32}$, $\|\rho_2, u\|_\lambda = (\frac{3}{4} - \frac{1}{2})^2 + (\frac{1}{4} - \frac{1}{2})^2 = \frac{1}{8}$, and $\|\rho_1, \rho_2\|_\lambda = (\frac{3}{8} - \frac{3}{4})^2 + (\frac{5}{8} - \frac{1}{4})^2 = \frac{9}{32}$, leading to $\mathsf{CH}_\lambda(\theta) = \frac{1}{3}(\frac{1}{2} - \frac{1}{32}) + \frac{2}{3}(\frac{1}{2} - \frac{1}{8}) + \frac{1}{3}\frac{2}{3}\frac{9}{32} = \frac{15}{32}$. □

We conclude by noting that this decomposition remains relevant even in the absence of panel data for, at least, two reasons. First, recent theoretical developments such as Dardanoni, Manzini, Mariotti, Petri, and Tyson (2023) show that under certain conditions the distribution and the behavior of types can be learnt from slightly enriched aggregate data. Second, in some situations, it may be the case that one of the components, say inter-personal heterogeneity, varies more across contexts or applications. In that case, a practitioner that is interested in studying this component of heterogeneity in a particular context but has only access to aggregate data may benefit from our decomposition result. The practitioner may leverage Proposition 1 to first estimate total heterogeneity from the aggregate data, determine a proxy for intra-personal heterogeneity from existing empirical findings and then use Proposition 3 to obtain an estimate of inter-personal heterogeneity.

5.4. **Mixing populations.** The decomposition obtained in Proposition 3 can be used in the evaluation of heterogeneity of the combination of any two populations θ and θ'. 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 $1-\alpha$. The next corollary follows directly from Proposition 3.

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

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

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.

A particular case of interest is that of the tremble model, where a population θ is mixed with uniform choices. Here, since the heterogeneity of uniform choices is higher than that of any other population, the mixing with uniform choices 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 2. *For every* $\alpha \in [0,1]$, $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1;u]) = \beta_\lambda - \alpha^2\|\rho_\theta, u\|_\lambda$.

Example 1 (continued). Let θ' be the population obtained by mixing the original population θ with $[1;u]$ with weights α and $1-\alpha$. That is, $\theta' = \alpha\theta + (1-\alpha)[1;u] = [\frac{\alpha}{3}, \frac{2\alpha}{3}, 1-\alpha; \rho_1, \rho_2, u]$. 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 Corollary 2, is $\frac{1}{2} - \alpha^2\frac{1}{32}$, a value that increases with the trembling weight $1-\alpha$. □

6. A characterization

We now provide a characterization of our measure of behavioral heterogeneity. Since our approach entails a family of measures parameterized by λ, we give three axioms, the combination of which is satisfied by every parameterization in our family, and only by them. We introduce each axiom in relation to a non-null generic heterogeneity function $\mathsf{H} : \Theta \to \mathbb{R}_+$, which assigns a level of heterogeneity to any possible population, such that $\mathsf{H}(\theta) = 0$ whenever θ is formed by a unique deterministic type, i.e. $\theta = [1;d]$. It is apparent that these populations have no variation whatsoever, and hence our basic assumption.

Our first axiom, Reduction, is the formalization of the ideas discussed in Section 5.1 regarding aggregate data.

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

For our next axiom, let Θ^D denote the set of populations with no intra-personal heterogeneity, that is, where all types in the population are deterministic. In line with the discussion in Section 5.4, we study the possibility of decomposing the heterogeneity of a deterministic population $\theta \in \Theta^D$ as the aggregation of sub-populations. In particular, consider hypothetical sub-populations each formed exclusively by two different deterministic types, with weights in proportion to their masses in the original population, i.e., sub-populations of the form $[\frac{\theta_i}{\theta_i+\theta_j}, \frac{\theta_j}{\theta_i+\theta_j}; d_i, d_j]$. 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 $\theta\in\Theta^D$, $\mathsf{H}(\theta)=\sum_{i<j}(\theta_i+\theta_j)^2\ \mathsf{H}([\frac{\theta_i}{\theta_i+\theta_j},\frac{\theta_j}{\theta_i+\theta_j};d_i,d_j])$.

Finally, we discuss a monotonicity property involving only couples, that is, populations composed exclusively of two equally-weighted deterministic types $[\frac{1}{2},\frac{1}{2};d_1,d_2]$. Consider collections of couples $C=\{[\frac{1}{2},\frac{1}{2};d_1^n,d_2^n]\}_{n=1}^N$. Now, take 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 $\overline{\mathsf{H}}(C)=\frac{\sum_n \mathsf{H}([\frac{1}{2},\frac{1}{2};d_1^n,d_2^n])}{N}$ 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 $\overline{\mathsf{H}}(C)\geq\overline{\mathsf{H}}(C')$.

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 $[1,\rho_\theta]$, and also equal to that of any deterministic population θ' with the same representative agent. Next, by Reduction and Decomposition, the heterogeneity of θ' can be broken down into the aggregation of the heterogeneities of all its couples, with weights derived from the masses of each deterministic type in θ'. Monotonicity permits to decompose further the former into couples that only differ in one menu, which can be seen as the contribution of each menu to total heterogeneity, i.e. $\lambda(A)$.

7. Random Utility Models

The use of particular stochastic choice models may facilitate the analysis. The canonical model in stochastic choice is the random utility model (RUM), and hence here we illustrate by focusing on the application of our measure of heterogeneity over populations of types described as RUMs. We show how, in this setting, the measure allows for an additional matrix representation, respects intuitive comparative statics of intra-personal heterogeneity and is characterized by the same set of three properties. Moreover, in Section 8 we comment that an additional benefit of specifying particular choice models is that they facilitate the structural estimation of individual stochastic choice functions even when only a single observation in some menus is available.

Let $\mathcal{P}$ denote the collection of all linear orders over X, which we call preferences. Consider a probability distribution ψ on $\mathcal{P}$. When choosing from menu $A \in \mathcal{A}$, each preference $P \in \mathcal{P}$ is realized with probability $\psi(P)$ and maximized. Denoting by $m(A, P)$ the maximal alternative in menu A according to preference P, the RUM probability that alternative a is selected in menu A is equal to:

$$\rho_\psi(a, A) = \sum_P \psi(P) \cdot \mathbb{1}_{[a=m(A,P)]}.$$

7.1. **A matrix representation.** Notice that in the present setting, a couple is simply $[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]$, where d_P denotes the deterministic type assigning probability one to the maximizer of preference P. 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 types differ in their choices. It is important to stress that this matrix is independent of the specific distribution over the types, and hence independent of the population, since it is characterized by the choice disagreements between preferences, weighted by 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.

Denote the probability distribution over $\mathcal{P}$ of the representative agent of θ by $\psi_\theta = \sum_i \theta_i \psi_i$. Proposition 4 shows that the choice heterogeneity of any population can be seen as an inner product involving its representative agent and matrix $\mathcal{C}_\lambda$.[5]

Proposition 4. $\mathsf{CH}_\lambda(\theta) = \psi_\theta \mathcal{C}_\lambda \psi_\theta^\top$.

Example 2.[6] 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

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

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

matrix reporting the heterogeneity of couples is

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

Consider now population $\theta = [1-\epsilon, \epsilon; d_{xyz}, u]$. The representative agent of the population is $\psi_\theta = (1 - \frac{5}{6}\epsilon, \frac{\epsilon}{6}, \ldots, \frac{\epsilon}{6})$. Hence, $\mathsf{CH}_{\bar{\lambda}}(\theta) = \psi_\theta \mathcal{C}_{\bar{\lambda}} \psi_\theta^\top = \frac{13\epsilon}{24}(2-\epsilon)$. $\square$

7.2. **Comparative statics: Intra-personal heterogeneity.** Proposition 3 shows that we can study the intra-personal heterogeneity of a type ρ_ψ by way of the λ-Euclidean distance between the behavior of the type and uniform choices, $\|\rho_\psi, u\|_\lambda$.[7] In order to investigate further the structure of intra-personal heterogeneity, we use a particular class of types, 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 type 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 types is the Luce model, as well as many of its generalizations such as the additive perturbed utility model of Fudenberg, Iijima and Strzalecki (2015).

Given two P-central types, ρ_ψ and $\rho_{\psi'}$, we say that the latter is a decentralization of the former if there exist $\epsilon > 0$ and preferences Q, Q' such that: (i) $\rho_{\psi'} = \rho_\psi - \epsilon d_Q + \epsilon d_{Q'}$ and (ii) Q' is farther away from P than Q is, i.e., xPy and $xQ'y$ imply xQy. That is, the second type is obtained from the first by shifting mass from preference Q to preference Q', 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.

[7] 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). Note also that u can be understood as the RUM assigning equal weight to every preference.

Formally, we say that $\rho_{\psi'}$ is a sequential decentralization of ρ_ψ whenever there is a sequence of decentralizations connecting ρ_ψ and $\rho_{\psi'}$.[8]

Proposition 5. *If $\rho_{\psi'}$ is a sequential decentralization of ρ_ψ, $\|\rho_\psi, u\|_\lambda \geq \|\rho_{\psi'}, u\|_\lambda$.*

Proposition 5 establishes some intuitive comparative statics on intra-personal heterogeneity for P-central types. 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. Formally, a Luce model can be described by means of a strictly positive real value function v, such that the choice probability of x in menu A is $\rho_v(x, A) = \frac{v(x)}{\sum_{y \in A} v(y)}$. Without loss of generality, we can normalize v to satisfy $\sum_{x \in X} v(x) = 1$. We can now establish the following result.

Proposition 6. *Suppose that $v(x_1) \geq \cdots \geq v(x_n)$ and $v'(x_1) \geq \cdots \geq v'(x_n)$. If $\frac{v'(x_j)}{v'(x_i)} \geq \frac{v(x_j)}{v(x_i)}$ for every $i < j$, $\|\rho_v, u\|_\lambda \geq \|\rho_{v'}, u\|_\lambda$.*

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

Example 3. Consider two Luce models with $v_1 = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$ and $v_2 = (\frac{5}{9}, \frac{3}{9}, \frac{2}{9})$. Since the monotone likelihood ratio holds for v_1 and v_2, Proposition 6 implies that $\|\rho_{v_1}, u\|_\lambda \geq \|\rho_{v_2}, u\|_\lambda$.[9] Now, consider the mixed-logit population given by $\theta = [\frac{4}{11}, \frac{7}{11}; \rho_{v_1}, \rho_{v_2}]$. Note that the representative agent of a mixed-logit population is not necessarily Luce, but must be a RUM. Since it can be seen that ρ_{v_2} is a decentralization of ρ_θ, which in turn is a decentralization of ρ_{v_1}, we can apply Proposition 5 implying that $\|\rho_\theta, u\|_\lambda \in [\|\rho_{v_2}, u\|_\lambda, \|\rho_{v_1}, u\|_\lambda]$.[10] $\square$

7.3. **A Characterization.** The characterization of CH_λ in the RUM framework follows very closely the general one obtained for the space of stochastic choice functions. The main difficulty in the present context is to obtain the contribution to heterogeneity of each menu A, because with RUMs it is impossible to find a couple that differs over a given menu A only. Hence, we need to expand the proof to identify two collections of couples for which the Δ-vectors differ only over menu A. We can then define the

[8] 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.

[9] Note that $\|\rho_{v_1}, u\|_\lambda = .1$ and $\|\rho_{v_2}, u\|_\lambda = .07$, consistent with the claim.

[10] Notice that $\|\rho_\theta, u\|_\lambda = .08$.

contribution to heterogeneity of menu A on the basis of the difference in heterogeneity between these two collections.

Theorem 2. *In the space of RUMs,* H *satisfies Reduction, Decomposition and Monotonicity if and only if there exists a probability distribution* λ *on* $\mathcal{A}$ *and* $k > 0$ *such that* $\mathsf{H} = k \cdot \mathsf{CH}_\lambda$.

8. Discussion

Our framework of behavioral heterogeneity is built upon the stochastic choices of a number of types, with θ_i describing the preponderance of type i in the population. This is clearly a simplification of real-life situations, where, typically, we would like to consider different type distributions for different menus. For instance, different individuals may have different probabilities of facing different menus. Note that this can be readily incorporated into the analysis by considering that population θ is described not by a unique distribution of types $\{\theta_1, \ldots, \theta_m\}$, but by menu-dependent distributions of types $\{\theta_1^A, \ldots, \theta_{m(A)}^A\}$. After sampling menu A with probability $\lambda(A)$, the measure would sample two individuals from the corresponding menu distribution of types and compare their choices. The basic structure of our measure remains unchanged.

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 some aspects related to the empirical implementation of our measure of choice heterogeneity. First, we want to emphasize again that our results are relevant even in the presence of aggregate data alone. This is because our overall measure can be directly implemented with such aggregate data and, as mentioned in Section 5.3, disaggregate population data may be recovered from sufficiently rich aggregate data following the techniques in Dardanoni, Manzini, Mariotti, Petri, and Tyson (2023).

Second, the ideal scenario entails datasets with stochastic choice data per type. The analysis is straightforward if types are defined as sub-groups of individuals, such as those given by age groups, gender, etc. When the focus is on individuals rather than types, stochastic choice datasets are more demanding, but there are relevant examples in the literature (see, e.g., Agranov and Ortoleva (2017), Alós-Ferrer and Garagnani (2021), and the papers cited therein).

Finally, notice that even when the practitioner lacks stochastic data at the individual level, she may recover the stochastic choice function of the individual from data by performing an estimation exercise for a particular structural model. In this vein, a series of papers propose 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), Dean, Ravindran, and Stoye (2022), de Clippel and Rozen (2022), Jagelka (2023), and Kocourek, Steiner, and Stewart (2023)).

Appendix A. Proofs

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

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

$$= \sum_A \lambda(A) \sum_i \sum_a \theta_i \rho_i(a, A) \sum_j \sum_b \theta_j \rho_j(b, A) \cdot \mathbb{1}_{[a \neq b]}$$

$$= \sum_A \lambda(A) \sum_a \rho_\theta(a, A) \sum_b \rho_\theta(b, A) \cdot \mathbb{1}_{[a \neq b]}$$

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

■

Proof of Proposition 2: We start by proving a series of useful claims. The first is that, conditional on having sampled the ordered pair of types (ρ, ρ'), the probability that a random choice from ρ disagrees with a random choice from ρ', over a random menu, can be written as:

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

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

To prove the claim, suppose that we have sampled the ordered pair of types (ρ,ρ'). Conditional heterogeneity is $\sum_A \lambda(A)\sum_a \rho(a,A)(1-\rho'(a,A))$, or equivalently

$$\sum_A \lambda(A)\sum_a [\rho(a,A)(1-\rho(a,A)) + \rho(a,A)(\rho(a,A)-\rho'(a,A))].$$

By similar reasoning, conditional heterogeneity is also equal to

$$\sum_A \lambda(A)\sum_a [\rho'(a,A)(1-\rho'(a,A)) + \rho'(a,A)(\rho'(a,A)-\rho(a,A))].$$

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

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

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

Second, we claim that for every population $\theta \in \Theta$, $\mathsf{CH}_\lambda(\theta) = \sum_i \theta_i \mathsf{CH}_\lambda([1;\rho_i]) + \sum_i \theta_i \sum_{i<j} \theta_j \; \|\rho_i,\rho_j\|_\lambda$. To see this, notice that $\mathsf{CH}_\lambda(\theta)$ is simply the aggregation of conditional heterogeneities across all possible ordered pairs of types weighted by their corresponding sampling probabilities. Hence, we proceed by aggregating the expression given above. Since every type i appears as the first type in the sampling with probability θ_i and again as the second type in the sampling with probability θ_i, the aggregation of conditional heterogeneities creates the value $\sum_i \theta_i \mathsf{CH}_\lambda([1;\rho_i])$. Given any pair $i<j$, these two types 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 $\sum_i \theta_i \sum_{i<j}\theta_j \; \|\rho_i,\rho'_j\|_\lambda$, thus proving the claim.

Third, we claim that for any behavior ρ, $\mathsf{CH}_\lambda([1;\rho]) = \beta_\lambda - \|\rho,u\|_\lambda$ holds. To see this, consider the population $\theta = [\frac{1}{2},\frac{1}{2};\rho,u]$. From the previous claim, $\mathsf{CH}_\lambda(\theta) = \frac{1}{2}\mathsf{CH}_\lambda([1;\rho]) + \frac{1}{2}\mathsf{CH}_\lambda([1;u]) + \frac{1}{4}\|\rho,u\|_\lambda$. Now, notice that, since one of the types involved is uniform, direct computation of the heterogeneity of θ yields $\mathsf{CH}_\lambda(\theta) = \frac{1}{4}\mathsf{CH}_\lambda([1;\rho]) + \frac{3}{4}\beta_\lambda$. By putting these two expressions together, we obtain:

$$\mathsf{CH}_\lambda([1;\rho]) = 3\beta_\lambda - 2\mathsf{CH}_\lambda([1;u]) - \|\rho,u\|_\lambda = 3\beta_\lambda - 2\beta_\lambda - \|\rho,u\|_\lambda = \beta_\lambda - \|\rho,u\|_\lambda.$$

Now, to prove the statement, note that Proposition 1 guarantees that $\mathsf{CH}_\lambda(\theta) = \mathsf{CH}_\lambda([1;\rho_\theta])$, and by the third claim $\mathsf{CH}_\lambda(\theta) = \beta_\lambda - \|\rho_\theta, u\|_\lambda$ holds. Finally, notice that $\max_\rho \|\rho, u\|_\lambda$ will be achieved by any deterministic type, 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 3: The proof follows directly from the second and third claims in the proof of Proposition 2. ∎

Proof of Corollary 1: The proof follows the same structure than the proof of Proposition 2. ∎

Proof of Corollary 2: From Corollary 1, $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1;u]) = \alpha\mathsf{CH}_\lambda(\theta) + (1-\alpha)\mathsf{CH}_\lambda([1;u]) + \alpha(1-\alpha)\|\rho_\theta, u\|_\lambda$, and Proposition 2 leads to $\mathsf{CH}_\lambda(\alpha\theta + (1-\alpha)[1;u]) = \alpha(\beta_\lambda - \|\rho_\theta, u\|_\lambda) + (1-\alpha)\beta_\lambda + \alpha(1-\alpha)\|\rho_\theta, u\|_\lambda = \beta_\lambda - \alpha^2\|\rho_\theta, u\|_\lambda$. ∎

Proof of Theorem 1: We start with the necessity part. The necessity of Reduction is shown in Proposition 1. For Decomposition, note that the probability that a deterministic type makes two different choices is zero, and hence the heterogeneity of $\theta \in \Theta^D$ can be written as

$$\begin{aligned}
\mathsf{CH}_\lambda(\theta) &= \sum_A \lambda(A) \sum_i \theta_i \sum_j \theta_j \sum_a d_i(a,A)(1-d_j(a,A)) \\
&= \sum_{i<j} (\theta_i+\theta_j)^2 \sum_A \lambda(A) \frac{2\theta_i\theta_j}{(\theta_i+\theta_j)^2} \mathbb{I}_{[d_i(A)\neq d_j(A)]} \\
&= \sum_{i<j} (\theta_i+\theta_j)^2\, \mathsf{CH}_\lambda\left(\left[\frac{\theta_i}{\theta_i+\theta_j}, \frac{\theta_j}{\theta_i+\theta_j}; d_i, d_j\right]\right).
\end{aligned}$$

For Monotonicity, note that the average heterogeneity of $C = \{[\frac{1}{2}, \frac{1}{2}; d_1^n, d_2^n]\}_{n=1}^N$ is:

$$\begin{aligned}
\overline{\mathsf{CH}}_\lambda(C) &= \frac{1}{N}\sum_n \sum_A \lambda(A)\frac{1}{2}\cdot \mathbb{I}_{[d_1^n(A)\neq d_2^n(A)]} = \frac{1}{2N}\sum_A \lambda(A)\sum_n \mathbb{I}_{[d_1^n(A)\neq d_2^n(A)]} \\
&= \frac{1}{2N}\sum_A \lambda(A)\Delta_A(C).
\end{aligned}$$

Given that λ is a positive-valued function, the necessity of Monotonicity follows. Finally, it is also immediate that $\mathsf{CH}_\lambda(\theta) = 0$ whenever $\theta = [1;d]$ and that the measure is non-null, as required by our basic assumptions over the heterogeneity map.

We now prove the sufficiency part. For every menu $A \in \mathcal{A}$, define

$$\tau(A) = \mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_1^A, d_2^A]),$$

where d_1^A and d_2^A are any two deterministic types such that coincide on every menu different than A, but differ in menu A. Notice that Monotonicity implies that $\tau(A)$ is independent of the selected pair of deterministic types.

We now show that for every pair of deterministic types $d_1, d_2 \in \Theta^D$, $\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_1, d_2]) = \sum_A \tau(A)\mathbb{1}_{[d_1(A) \neq d_2(A)]}$. Let n be the cardinality of the set $\{A : d_1(A) \neq d_2(A)\}$. If $n = 0$, we know by assumption that $\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_1, d_2]) = 0$, as desired. Then, let us assume that $n > 0$. For every menu A such that $d_1(A) \neq d_2(A)$, consider a couple of the form $[\frac{1}{2}, \frac{1}{2}; d_1^A, d_2^A]$ where d_1^A, d_2^A are two deterministic types differing only over menu A. Consider then two collections of couples C_1 and C_2, such that: (i) C_1 is formed by $[\frac{1}{2}, \frac{1}{2}; d_1, d_2]$ and $n-1$ copies of $[1; d]$ for any $d \in \Theta^D$, and (ii) C_2 is formed by the n couples defined above $[\frac{1}{2}, \frac{1}{2}; d_1^A, d_2^A]$. These are two equally-sized collections of couples, such that $\Delta_A(C_1) = \Delta_A(C_2)$ for every A. Monotonicity, the normalization, and the definition of τ imply the claim.

Now, for any pair of deterministic types d_1, d_2, we show that $\mathsf{H}([1-\gamma, \gamma; d_1, d_2]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_1, d_2])$ for every constant $\gamma \in [0, 1]$. Consider any two values $\gamma_1, \gamma_2 \in [0, 1]$ and the mixing of populations $[1-\gamma_1, \gamma_1; d_1, d_2]$ and $[1-\gamma_2, \gamma_2; d_1, d_2]$ with weights $\frac{\gamma_2}{\gamma_1+\gamma_2}$ and $\frac{\gamma_1}{\gamma_1+\gamma_2}$. That is, $\theta' = [\frac{\gamma_2}{\gamma_1+\gamma_2}(1-\gamma_1), \frac{\gamma_1}{\gamma_1+\gamma_2}(1-\gamma_2), \frac{\gamma_2}{\gamma_1+\gamma_2}\gamma_1, \frac{\gamma_1}{\gamma_1+\gamma_2}\gamma_2; d_1, d_1, d_2, d_2]$. Since $\theta' \in \Theta^D$, the application of Decomposition, together with the fact that homogeneous and deterministic populations have zero heterogeneity, leads to

$$\mathsf{H}(\theta') = 2[(\frac{\gamma_2}{\gamma_1+\gamma_2})^2\,\mathsf{H}([1-\gamma_1, \gamma_1; d_1, d_2) + (\frac{\gamma_1}{\gamma_1+\gamma_2})^2\,\mathsf{H}([1-\gamma_2, \gamma_2; d_1, d_2)].$$

Since we have $\frac{\gamma_2}{\gamma_1+\gamma_2}(1-\gamma_1) + \frac{\gamma_1}{\gamma_1+\gamma_2}(1-\gamma_2) = \frac{\gamma_1+\gamma_2-2\gamma_1\gamma_2}{\gamma_1+\gamma_2} = \gamma_3$, Reduction guarantees that the heterogeneity of population $[1-\gamma_3, \gamma_3; d_1, d_2]$ must be equivalent to that of θ', leading to

$$\mathsf{H}([1-\gamma_3, \gamma_3; d_1, d_2) = 2[(\frac{\gamma_2}{\gamma_1+\gamma_2})^2\mathsf{H}([1-\gamma_1, \gamma_1; d_1, d_2) + (\frac{\gamma_1}{\gamma_1+\gamma_2})^2\mathsf{H}([1-\gamma_2, \gamma_2; d_1, d_2)].$$

Notice that $4(1-\gamma_3)\gamma_3 = 2((\frac{\gamma_2}{\gamma_1+\gamma_2})^2 4\gamma_1(1-\gamma_1) + (\frac{\gamma_1}{\gamma_1+\gamma_2})^2 4\gamma_2(1-\gamma_2))$, and given that the former must be true for every pair of values of γ_1 and γ_2, it must be $\mathsf{H}([1-\gamma, \gamma; d_1, d_2]) = 4\gamma(1-\gamma)\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_1, d_2])$.

We now show that for every $\theta \in \Theta^D$, $\mathsf{H}(\theta) = \sum_{i<j} 4\,\theta_i\theta_j\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_i, d_j])$. Using Decomposition and the former step,

$$
\begin{aligned}
\mathsf{H}(\theta) &= \sum_{i<j} (\theta_i + \theta_j)^2 \, \mathsf{H}\left(\left[\frac{\theta_i}{\theta_i + \theta_j}, \frac{\theta_j}{\theta_i + \theta_j}; d_i, d_j \right] \right) \\
&= \sum_{i<j} (\theta_i + \theta_j)^2 \cdot 4 \frac{\theta_i}{\theta_i + \theta_j} \frac{\theta_j}{\theta_i + \theta_j} \mathsf{H}\left(\left[\frac{1}{2}, \frac{1}{2}; d_i, d_j \right] \right) \\
&= \sum_{i<j} 4 \theta_i \theta_j \mathsf{H}\left(\left[\frac{1}{2}, \frac{1}{2}; d_i, d_j \right] \right).
\end{aligned}
$$

Finally, we show that $\mathsf{H} = k \cdot \mathsf{CH}_\lambda$ for some $k > 0$. Consider any population θ and any deterministic population $\theta' \in \Theta^D$ such that $\rho_\theta = \rho_{\theta'}$. We have shown that $\mathsf{H}(\theta') = \sum_{i<j} 4\, \theta'_i \theta'_j \mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_i, d_j])$. Also, we have shown that $\mathsf{H}(\theta') = \sum_{i<j} 4\, \theta'_i \theta'_j \sum_{A: d_i(A) \neq d_j(A)} \tau(A)$. Since the measure is non-null, it cannot be that $\tau(A) = 0$ for every A, and we can define the probability distribution $\lambda(A) = \frac{\tau(A)}{\sum_A \tau(A)}$. We can then rewrite the expression above as $\mathsf{H}(\theta') = k \sum_A \lambda(A) \sum_i \theta'_i \sum_j \theta'_j \mathbb{1}_{[d_i(A) \neq d_j(A)]}$, which, given the fact that θ' is deterministic, coincides with $\mathsf{CH}_\lambda(\theta')$. Now since the representative agent of θ' coincides with that of θ, Reduction (and the fact that CH_λ satisfies this property) guarantees that $\mathsf{H}(\theta) = \mathsf{H}(\theta') = \mathsf{CH}_\lambda(\theta') = \mathsf{CH}_\lambda(\theta)$. This concludes the proof. ■

Proof of Proposition 4: The proof follows from the proof of Theorem 2. ■

Proof of Proposition 5: Suppose that $\rho_{\psi'}$ is a sequential decentralization of ρ_ψ. By definition, there exists a sequence $\{\rho_{\psi^j}\}_{j=1}^J$ of types such that $\rho_{\psi^1} = \rho_\psi$, $\rho_{\psi^J} = \rho_{\psi'}$, and ρ_{ψ^j} is a decentralization of $\rho_{\psi^{j-1}}$ for $j = 2, \ldots, J$, with the central preference denoted as P. At each stage j, mass $\epsilon^j > 0$ shifts from preference Q^j to another preference Q'^j, i.e., $\rho_{\psi^{j+1}} = \rho_{\psi^j} - \epsilon^j d_{Q^j} + \epsilon^j d_{Q'^j}$. 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^j and Q'^j differ in their ranking of only two alternatives, with $x^j P y^j, x^j Q^j y^j$ and $y^j Q'^j x^j$.

First, consider any menu A that does not contain either x^j or y^j or such that $m(A, Q^j) \neq x_j$. Preferences Q^j and Q'^j have the same maximizer over such a menu and hence, it is evident that $\rho_{\psi^{j+1}}(A) = \rho_{\psi^j}(A)$, 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 $x^j = m(A, Q^j)$. 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 $[1; \rho_{\psi^j}]$ within menu A is equal to $1 - \sum_{z \in A}(\rho_{\psi^j}(z, A))^2$, the transfer must increase the heterogeneity of menu A. Additivity across menus guarantees that $\mathsf{CH}_\lambda([1; \rho_{\psi^{j+1}}]) \geq \mathsf{CH}_\lambda([1; \rho_{\psi^j}])$. The recursive application of this argument over the sequence of types together with Proposition 2 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 $v(y_1) \geq \cdots \geq v(y_K)$ and $v'(y_1) \geq \cdots \geq v'(y_K)$. First, notice that the assumption guarantees that $\frac{v'(y_s)}{v'(y_t)} \geq \frac{v(y_s)}{v(y_t)}$ for every $s > t$ and, hence, $\frac{\rho_{v'}(y_s,A)}{\rho_{v'}(y_t,A)} = \frac{\frac{v'(y_s)}{\sum_{k=1}^K v'(y_k)}}{\frac{v'(y_t)}{\sum_{k=1}^K v'(y_k)}} \geq \frac{\frac{v(y_s)}{\sum_{k=1}^K v(y_k)}}{\frac{v(y_t)}{\sum_{k=1}^K v(y_k)}} = \frac{\rho_v(y_s,A)}{\rho_v(y_t,A)}$. 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 $\rho_v(y_t, A) \geq \rho_{v'}(y_t, A)$ if and only if $t \leq T$. Since $\sum_{k=1}^K \rho_v(y_k, A) = \sum_{k=1}^K \rho_{v'}(y_k, A) = 1$, the distribution $\{\rho_{v'}(y_k, A)\}_{k=1}^K$ second-order stochastically dominates the distribution $\{\rho_v(y_k, A)\}_{k=1}^K$. The strict convexity of the quadratic function guarantees that $\frac{\sum_{k=1}^K(\rho_v(y_k,A))^2}{K} \geq \frac{\sum_{k=1}^K(\rho_{v'}(y_k,A))^2}{K}$, or equivalently $\sum_{k=1}^K(\rho_v(y_k, A))^2 \geq \sum_{k=1}^K(\rho_{v'}(y_k, A))^2$. Conditional on menu $A \in \mathcal{A}$, we can write intra-personal heterogeneity as 1 minus the previous sums of squares, and the additivity of intra-personal heterogeneity across menus concludes the proof. ■

Proof of Theorem 2: The proof follows the strategy used in the proof of Theorem 1. Here we discuss the main differences only, that arise from the fact that in the RUM framework the implications on behavioral heterogeneity of a couple $[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]$ are more intricate, in the sense that they necessarily involve differences on a number of menus, not on just one menu.

First of all, we need to obtain $\tau(A)$ using a more elaborated method. 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 $(X \setminus B)\, P_B^A\, a\, P_B^A\, b\, P_B^A\, (B \setminus \{a, b\})$, and a preference Q_B^A obtained from P_B^A by swapping the position of alternatives a and b in P_B^A. We define

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

We need to show that this expression is independent of the selected pair of alternatives and collection of preferences. 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 $[\frac{1}{2}, \frac{1}{2}; d_{P_B^A}, d_{Q_B^A}]$ where $\{a, b\} \subset B \subseteq A$ is such that $(-1)^{|A|-|B|} = 1$, (ii) C_2^A is formed by all couples $[\frac{1}{2}, \frac{1}{2}; d_{P_B^A}, d_{Q_B^A}]$ where $\{a, b\} \subseteq B \subseteq A$ is such that $(-1)^{|A|-|B|} = -1$, (iii) $C_1'^A$ is the collection of all couples $[\frac{1}{2}, \frac{1}{2}; d_{P_{B'}'^A}, d_{Q_{B'}'^A}]$ where $\{a', b'\} \subseteq B' \subseteq A$ satisfies $(-1)^{|A|-|B'|} = 1$ and, finally (iv) $C_2'^A$ is formed by all couples $[\frac{1}{2}, \frac{1}{2}; d_{P_{B'}'^A}, d_{Q_{B'}'^A}]$ where $\{a', b'\} \subset B' \subseteq A$ 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

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

Second, we now obtain the heterogeneity of a couple $[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]$ on the basis of the sum of the τ contributions of all those menus where types d_P and d_Q differ. Formally, we now show that 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 $\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_P, d_Q]) = 0$, 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 construction of τ.

Consider the following two collections of couples: (i) $\bigcup_{A:m(A,P)\neq m(A,Q)} C_1^A$ and (ii) $\bigcup_{A:m(A,P)\neq m(A,Q)} C_2^A \cup \{[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]\}$. 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 $[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]$. Hence, if $n = 0$, select any preference R and add the population $[1; d_R] = [\frac{1}{2}, \frac{1}{2}; d_R, d_R]$ to (i). If $n > 1$, add $n-1$ copies of the population $[1; d_R] = [\frac{1}{2}, \frac{1}{2}; d_R, d_R]$ to (ii). In any case, we have defined two equally-sized collections of couples which we call, respectively, C and C'.

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 $[\frac{1}{2}, \frac{1}{2}; d_R, d_R]$ are irrelevant in this respect, and population $[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]$ is such that $\Delta_A(\{[\frac{1}{2}, \frac{1}{2}; d_P, d_Q]\}) = 1$ 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 $\mathsf{H}([\frac{1}{2}, \frac{1}{2}; d_R, d_R]) = 0$, we can apply Monotonicity to obtain

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

It then follows that

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

The rest of the proof follows the steps of Theorem 1. ■

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] Alós-Ferrer, C. and M. Garagnani (2021). "Choice consistency and strength of preference." *Economics Letters*, 198, 109672.

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

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

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

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

[9] Bertrand, M. and E. Kamenica (2023). "Coming Apart? Cultural Distances in the United States over Time." *American Economic Journal: Applied Economics*, forthcoming.

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

[11] Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and C.J. Tyson (2023). "Mixture Choice Data: Revealing Preferences and Cognition." *Journal of Political Economy*, 131(3), 687-715.

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

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

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

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

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

[17] Fudenberg, D., R. Iijima, and T. Strzalecki (2015). "Stochastic choice and revealed perturbed utility." *Econometrica*, 83(6), 2371-2409.

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

[19] 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.

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

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

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

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

[24] Kocourek, P., J. Steiner, and C. Stewart (2023). "Boundedly Rational Demand." Mimeo.

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

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

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

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

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

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