Mixture Choice Data: Revealing Preferences *and* Cognition*

Valentino Dardanoni, Paola Manzini, Marco Mariotti,

Henrik Petri, and Christopher J. Tyson†

May 28, 2022

Abstract

Mixture choice data consist of the joint distribution of choices of a group of agents from a collection of menus, comprising the implied stochastic choice function plus any cross-menu correlations. When agents are heterogeneous with respect to both preferences and other aspects of cognition, we show that these two determinants of behavior are identified simultaneously by suitable mixture choice data. We also demonstrate how this finding can be extended to allow for specialized assumptions about cognition, focusing on models of random satisficing thresholds and "quantal Fechnerian" choice.

J.E.L. classification codes: D01, D91.
Keywords: cognition, heterogeneity, identification, satisficing, stochastic choice.

---

*For valuable comments and suggestions on this paper we wish to thank Federico Echenique, Emel Filiz-Ozbay, Guillaume Fréchette, David Freeman, Yusufcan Masatlioglu, and Illia Pasichnichenko; as well as numerous seminar and conference audiences. Dardanoni thanks the MIUR for financial support through PRIN Grant 2017KZZLYP. Manzini and Mariotti thank the Leverhulme Foundation for support through Research Project Grant 2019–143.

†Dardanoni: Università degli Studi di Palermo.
Manzini: University of Bristol and IZA.
Mariotti: Queen Mary University of London and Deakin University.
Petri: University of Bath.
Tyson: Queen Mary University of London.

1 Introduction

Samuelson [35], Arrow [8], and their successors explored the behavioral consequences of preference-based choice, showing how unobserved tastes could be recovered from observed actions. While this research program succeeded in establishing more solid foundations for traditional models of utility maximization, it did not capture two fundamental aspects of real-world decision making that later researchers have tried to incorporate: Firstly, data available in practice often record the *aggregate* behavior of a population of agents with *heterogeneous* tastes. Secondly, choices may be influenced by a variety of cognitive factors other than binary preference comparisons. Individually and in combination, these issues complicate the task of using data to uncover the unobserved components of a model, thus posing a challenge to the elaboration of the revealed preference program.

In response to the first issue, a strand of research originating with Luce [27] and Block and Marschak [10] focuses on the rationalization of stochastic choice data, interpreted as the aggregate behavior of a population of heterogeneous individuals. This work has led to a highly productive interplay with econometrics, as well as to the development of discrete choice models that have become essential tools in labor economics, industrial organization, and other fields.[1] However, allowing heterogeneous tastes raises immediate difficulties for identification: The random utility model (RUM), which incorporates heterogeneity in the natural way by positing a probability distribution over utility functions, is known to have poor identification properties. Fishburn [18] has demonstrated that it is even possible for stochastic choices from all possible menus to be generated by two distinct RUMs with disjoint support.[2] These observations have led to the study of restrictions on the random utility framework that improve identification, such as those captured in the logit model of Luce [27] and McFadden [31], or the single-crossing RUM of Apesteguia et al. [7].

[1] Recent theoretical contributions that advance this agenda include Apesteguia et al. [6] and Kovach and Tserenjigmid [22].

[2] In Fishburn's example, one RUM assigns equal probability to $w \succ x \succ y \succ z$ and $x \succ w \succ z \succ y$, while another assigns equal probability to $w \succ x \succ z \succ y$ and $x \succ w \succ y \succ z$. These models yield the same stochastic choices over any subset of the four alternatives. (See also Turansick [39].)

Turning to the second issue mentioned above, it has long been acknowledged that preferences are not the sole determinant of choice behavior. The agent's working memory, cognitive load, mood or affect, framing of the situation, ingrained biases and heuristics, level of attention, and general intelligence can all be shown to influence decision making in ways that are difficult to reconcile with the preference maximization paradigm. Moreover, there is no compelling reason to expect a population of agents to be more homogeneous in respect of these cognitive traits and abilities than in respect of their tastes. Ideally, we would like to allow heterogeneity on both dimensions—and even correlation between them—while continuing to work with stochastic choice data. But many existing models with elaborated cognition either inhabit a deterministic setting (e.g., Dutta and Horan [15] and Lleras et al. [26]) or assume that tastes are uniform across the population (e.g., Aguiar et al. [2] and Manzini and Mariotti [28]).

Of course, it is straightforward to build a model that allows "double heterogeneity" of the desired sort. The difficulty here lies in constructing a framework that can combine this flexibility with attractive identification properties, enabling us to distinguish preferences and cognition and to pin down the parameters of each component. Indeed, mixing taste and cognitive heterogeneity tends to exacerbate the identification problems arising from each dimension separately. For instance, models of these two components that are identified in isolation may not be identified when they are merged.

For a concrete illustration of this point, consider the stochastic choice function (SCF) in Table 1A.[3] Although the random utility model in general is not identified, the choice data shown here do have a unique RUM representation.[4] Alternatively, fixing the preferences $x \succ y \succ z$, we can explain the same data using a model of satisficing behavior from the class defined below in Section 4.2. Here a threshold alternative $\tilde{x}$ is drawn from the menu

[3]In Table 1 and throughout, note the multiplicative notation for enumerated sets of alternatives. We use this notation for menus to avoid a proliferation of $\{\cdot\}$ symbols, but not for other kinds of sets as this could create ambiguity.

[4]The SCF in Table 1A admits the unique taste distribution $\Pr[x \succ y \succ z] = 0.2$, $\Pr[x \succ z \succ y] = 0.3$, $\Pr[y \succ x \succ z] = 0.1$, $\Pr[y \succ z \succ x] = 0.2$, $\Pr[z \succ x \succ y] = 0.1$, and $\Pr[z \succ y \succ x] = 0.1$. Note that this distribution satisfies Turansick's [39] conditions for RUM identification.

	[A.]			[B.]		
A	$\rho(x, A)$	$\rho(y, A)$	$\rho(z, A)$	$\tau_A(x)$	$\tau_A(y)$	$\tau_A(z)$
xy	0.6	0.4	0.0	0.2	0.8	—
xz	0.6	0.0	0.4	0.2	—	0.8
yz	0.0	0.5	0.5	—	0.0	1.0
xyz	0.5	0.3	0.2	0.2	0.2	0.6

Table 1: [A.] A stochastic choice function over the menus drawn from $X = xyz$. [B.] The corresponding threshold distributions for a satisficing representation of the choice data.

A according to the distribution τ_A shown in Table 1B. The satisficing agent considers all members of A that are weakly preferred to the threshold $\tilde{x}$ to be acceptable choices from this menu, and selects one of them uniformly. For instance, we have

$$\rho(x, xyz) = \frac{\overbrace{\tau_{xyz}(x)}^{\tilde{x}=x}}{1} + \frac{\overbrace{\tau_{xyz}(y)}^{\tilde{x}=y}}{2} + \frac{\overbrace{\tau_{xyz}(z)}^{\tilde{x}=z}}{3} = \frac{0.2}{1} + \frac{0.2}{2} + \frac{0.6}{3} = 0.5, \tag{1}$$

$\rho(y, xyz) = 0.2/2 + 0.6/3 = 0.3$, and $\rho(z, xyz) = 0.6/3 = 0.2$. Importantly, like the RUM representation of the same data, the satisficing representation in Table 1B is unique.[5]

Now suppose that we combine this type of satisficing with preference heterogeneity, assuming for purposes of this example that tastes and thresholds are drawn independently. The parameters of the resulting model are no longer uniquely determined by the data in Table 1A. On the one hand, we could have the taste distribution $\Pr[x \succ y \succ z] = 0.6$ and $\Pr[z \succ y \succ x] = 0.4$, together with the threshold distributions

$$\tau_{xy}(x) = 0.6,\ \tau_{xy}(y) = 0.4;\ \tau_{xz}(x) = 0.6,\ \tau_{xz}(z) = 0.4;\ \tau_{yz}(y) = 0.4,\ \tau_{yz}(z) = 0.6;$$
$$\tau_{xyz}(x) = 0.48,\ \tau_{xyz}(y) = 0.44,\ \tau_{xyz}(z) = 0.08. \tag{2}$$

[5]Incidentally, the tastes $x \succ y \succ z$ can be inferred from the data $\rho(x, xyz) > \rho(y, xyz) > \rho(z, xyz)$, and so need not be known in advance.

On the other hand, we could have $\Pr[x \succ y \succ z] = 0.8$ and $\Pr[z \succ y \succ x] = 0.2$, with

$$\tau_{xy}(x) = 0.4,\ \tau_{xy}(y) = 0.6;\ \tau_{xz}(x) = 0.4,\ \tau_{xz}(z) = 0.6;\ \tau_{yz}(y) = 0.2,\ \tau_{yz}(z) = 0.8;$$
$$\tau_{xyz}(x) = 0.3,\ \tau_{xyz}(y) = 0.4,\ \tau_{xyz}(z) = 0.3. \tag{3}$$

Here taste heterogeneity disrupts identification of the satisficing model, and vice versa, even though the SCF in question poses no difficulties for either model in isolation.

The above illustration shows that for models in which non-preference factors affect behavior, stochastic choice functions do not provide rich enough data for identification purposes under even the most basic forms of taste heterogeneity. To address this challenge, we posit the availability of "mixture choice" data, which consist of the *joint* distribution of choices of a population of agents from a finite collection of menus. For example, in an experiment where subjects face multiple decision problems, the researcher will know not only which option is selected from each menu separately—that is, the empirical stochastic choice function—but also the proportion of subjects who make a given combination of choices across all problems. On a larger scale, businesses such as supermarkets, credit card issuers, and online retailers can track the choices of individuals over time as menus change due to product availability and entry or exit from particular markets. Panel or focus group data will also have the mixture choice form when the menus faced by members are reliably recorded. Mixture choice data are thus in no way an exotic theoretical construct, and are in fact increasingly prevalent in the information-oriented economy.

Formally, a *mixture choice function* (MCF) is defined as a probability distribution over deterministic choice functions (i.e., maps from menus to choices), and can be interpreted as returning the share of the population that chooses in a particular way across the entire domain of problems in a dataset. As noted, an MCF contains richer information than the corresponding SCF, since it records correlations among the choices from different menus. We will find that knowledge of these cross-menu correlations is essential to achieve our goal of disentangling preferences from other factors influencing behavior.

In our framework, the mixture choice data are generated by a population of agents possessing individual "types." Mathematically, a type is simply the index for a particular SCF, and so the observed choice probabilities can be described as an *unobserved mixture* (over types) of *unobserved type-conditional stochastic choices* from each menu.[6] An agent's type thus captures the menu-independent features of the choice procedure governing his or her behavior, which may also be affected by menu-specific factors not encoded in the type. Under one natural interpretation, a type is a preference order over the alternatives that is common to all choice problems, but may be maximized imperfectly due to factors such as framing, mood, or attention. These menu-specific factors are realized independently across the menus in the dataset, while preferences are assumed to remain fixed. Alternatively, in a procedural model of choice the type could represent the agent's general intelligence, presumed to be stable across problems, and the residual type-conditional variation could capture the agent's cognitive load when facing different menus. Our results do not depend on any particular interpretation of types, leaving scope in applications for the modeler to define them appropriately to context.

A mixture choice dataset is exactly identified (i.e., point identified) if it has a unique decomposition into type-conditional SCFs. In this case the data will reveal each type's share of the population plus each type's distribution of choices when any menu-specific factors are realized. In Section 3, we offer a variety of results (Propositions 1–5) showing that a unique decomposition is guaranteed *generically*, provided the menus in our dataset are large and numerous enough relative to the number of types.[7]

The first message of these results is quantitative: As the size and number of the menus in our mixture choice dataset increases, the number of types that can be identified grows explosively. For instance, with choice data for all menus drawn from a set of five or more

[6] A more formal expression of the structure we impose on mixture choice data appears in Equation 4. For a review of finite mixture models in econometrics, see Compiani and Kitamura [13]. For a prominent example of their use in experiments, see Harrison and Rutström [19].

[7] These identification results rely on a powerful theorem from Allman et al. [4], adapted as Lemma 1.

alternatives, the number of identifiable types exceeds the number of possible preference orders by many orders of magnitude (Proposition 2). Enormous type spaces can also be accommodated using menus of uniform size drawn from a modestly proportioned set of alternatives (Proposition 3).

Our results may also be useful for experimental design. Suppose, for instance, that we wish to use a choice experiment to determine the taste distribution in a population of agents who are also influenced by non-preference factors. How many choice problems are needed? How do menus of different sizes contribute to achieving identification? If adding new menus or expanding existing menus is costly—perhaps due to subjects becoming fatigued or their cognitive imperfections being exacerbated—what is the most efficient way to construct the menus presented? Our findings are well suited to address questions such as these.

After investigating sufficient conditions for identification in the general mixture choice framework, we extend the analysis to two specialized settings in which types are preference orders and the type-conditional variation results from some form of bounded rationality. Here we describe models of random satisficing thresholds (Section 4.2) and of "quantal Fechnerian" choice (Section 4.3), each involving cognitive parameters that control the extent of departure from preference maximization. The objective is to elicit these "deep" parameters using our knowledge of the type-conditional SCFs, and to ensure that generic identification continues to hold in these specialized settings (Proposition 6). Notably, our approach allows for correlation between preference and non-preference factors affecting choice, as well as for menu-dependence of the deep parameters of a model.

We conclude this section with a brief survey of related work. Firstly, a small recent literature has made important progress on the problem of achieving identification when both preferences and cognitive characteristics are heterogeneous. In particular, Abaluck and Adams [1], Aguiar et al. [3], Barseghyan et al. [9], and Sovinsky Goeree [36] examine models in which—due to limited attention—the decision maker considers only a subset of the available alternatives. These contributions either obtain partial identification (i.e., set identification) or gain the leverage needed for exact identification via agent covariates (e.g., demographics), alternative covariates (e.g., advertising levels), or experimental control of cognition. On the other hand, Kashaev and Aguiar [21] avoid the use of covariates and study SCFs generated by the random attention model of Cattaneo et al. [12]. They too allow for heterogeneous preferences and derive partial identification results.

We aim to develop a methodology for exact identification that can be adapted to a range of models of decision making, while at the same time requiring only choice data. To achieve this, we posit a dataset richer than those used in the papers mentioned above. Specifically, we introduce the concept of a mixture choice function, simultaneously with and independently of the work of Filiz-Ozbay and Masatlioglu [17]. These authors study the relationship of mixture choice functions to the corresponding SCFs, proving a number of results about the representation of stochastic choice data using generalizations of the single-crossing RUM of Apesteguia et al. [7]. In contrast, we focus on identification of the structural parameters embedded in mixture choice data, which we take to be observable, and in this sense the two papers can be seen as complementary.[8]

Finally, the present paper builds upon Dardanoni et al. [14], where techniques from the statistical literature on tensor decompositions were used for identification purposes in the context of a population of decision makers with limited "consideration capacity." This framework is both more specialized than that of Sections 2–3 below and distinct from the models studied in Section 4. Moreover, our emphasis will be on menu variation, departing from Dardanoni et al. [14] and returning the tensor decomposition approach to the original proving ground of revealed preference theory.

[8]Another related contribution is that of Lin [25], who shows by example how observing the correlations between choices can help to achieve identification in a random *expected* utility model.

2 The general mixture choice model

Let X be a finite set of *alternatives*, with cardinality $n \geq 3$. A *menu* $A \subseteq X$ is any set of alternatives with $|A| \geq 2$; and $\mathcal{A}$ denotes a fixed collection of $K \geq 3$ such menus, which we will enumerate as $\langle A_1, \ldots, A_K \rangle$ when convenient. We refer to the set $\{A \subseteq X : |A| \geq 2\}$ as the *full collection* of menus.

A *choice function* is any rule $c : \mathcal{A} \to X$ such that $\forall A \in \mathcal{A}$ we have $c(A) \in A$, and $\mathbb{C}$ denotes the set of all choice functions. In this setting, a *stochastic choice function* (SCF) is a $\rho : X \times \mathcal{A} \mapsto [0, 1]$ such that for each $A \in \mathcal{A}$ we have $\sum_{x \in A} \rho(x, A) = 1$. We study the following more general notion of random choice, which permits statistical dependence of choice behavior across menus.

Definition 1. A *mixture choice function* (MCF) is a probability distribution over $\mathbb{C}$; i.e., a $\mu : \mathbb{C} \to [0, 1]$ such that $\sum_{c \in \mathbb{C}} \mu(c) = 1$. The MCF μ is said to be *menu independent* if $\forall c \in \mathbb{C}$ we have $\mu(c) = \prod_{A \in \mathcal{A}} \sum_{c' \in \mathbb{C} : c'(A) = c(A)} \mu(c')$.

Menu independence eliminates interaction across choice problems, returning us to the SCF framework. To see this, note that any SCF ρ induces the ("product") MCF defined by $\mu_\rho(c) = \prod_{A \in \mathcal{A}} \rho(c(A), A)$, and conversely any MCF μ can be used to construct the ("marginal") SCF $\rho_\mu(x, A) = \sum_{c \in \mathbb{C} : c(A) = x} \mu(c)$. The mapping $\rho \mapsto \mu_\rho$ is then a bijection from the set of all stochastic choice functions onto the set of menu-independent mixture choice functions, with inverse map $\mu \mapsto \rho_\mu$.[9]

Our objective is to use mixture choice functions to achieve simultaneous identification of preferential and cognitive factors affecting choice behavior, when agents are potentially heterogeneous on *both* dimensions. Specifically, we imagine a population of decision makers with *types* drawn from a finite set Θ. On the one hand, the type θ may encode the agent's preferences—either a full preference order over X or the value of a (discretized)

[9] The equivalence between SCFs and menu-independent MCFs relies on the latter being defined over the entire space $\mathbb{C}$, and may not hold on smaller spaces such as the class of rational choice functions. For an analysis of MCFs that possess structure of this sort, see Filiz-Ozbay and Masatlioglu [17].

numerical preference parameter such as a discount factor or a coefficient of risk aversion. On the other hand, the type could capture a cognitive trait such as a reference point or level of attention to the choice environment.

The type distribution is denoted by $\langle \pi(\theta) \rangle_{\theta \in \Theta}$, and each type θ has a stochastic choice function ρ^θ realized independently across the menus in the collection $\mathcal{A}$. We assume that the researcher observes the mixture choice function

$$\mu(c) = \sum_{\theta \in \Theta} \pi(\theta) \prod_{A \in \mathcal{A}} \rho^\theta(c(A), A). \tag{4}$$

In other words, the likelihood of choice function c is the probability-weighted sum of the likelihood that type θ chooses alternative $c(A)$ independently from each menu A.

Observe that the determination of $\mu(c)$ via Equation 4 involves two nested levels of stochasticity: We are uncertain about the pattern of choices that will arise both because the type θ is random and because, conditional on type, the choice function ρ^θ is random. For example, if types indicate the framing of the problem (Salant and Rubinstein [34]), then choices are stochastic both because we have imperfect knowledge of the frame itself and because other nondeterministic factors influence behavior, conditional on the frame. These factors could include preferences as well as other cognitive variables such as the decision maker's level of attention (Masatlioglu et al. [30]), mood or other affective state (Manzini and Mariotti [29]), or capacity to process information (Wilson [41]).[10]

Whatever the interpretation of θ, our primary assumption is that the agent's choice behavior is statistically independent across menus, conditional on type. Mathematically, this requires that the type-specific MCFs be products over A of the marginal probabilities $\rho^\theta(c(A), A)$; a property that is not in general inherited by the observed aggregate MCF.

[10]If each function ρ^θ were deterministic, then $\mu(c)$ would simply be the probability of the type θ (if any) whose behavior conformed to c. Here μ would reveal the type distribution directly, and identification of the model would not be an issue. In the example, each frame would yield a different pattern of choices for fixed preferences, and hence the frame distribution could be inferred from μ in the absence of preference variation. In contrast, Equation 4 allows frames and preferences to vary simultaneously, in a manner that confounds their effects but nevertheless preserves the possibility of identification from the data in μ.

[A.]

θ	$\rho^\theta(x,xy)$	$\rho^\theta(y,xy)$	$\rho^\theta(x,xz)$	$\rho^\theta(z,xz)$	$\rho^\theta(y,yz)$	$\rho^\theta(z,yz)$
θ_1	3/5	2/5	3/4	1/4	2/3	1/3
θ_2	2/3	1/3	2/5	3/5	1/4	3/4

[B.]

	$c(xy)=x$ $c(yz)=y$	$c(xy)=x$ $c(yz)=z$	$c(xy)=y$ $c(yz)=y$	$c(xy)=y$ $c(yz)=z$
$c(xz)=x$	22/120	21/120	14/120	12/120
$c(xz)=z$	12/120	21/120	7/120	11/120

[C.]

$\rho_\mu(x,xy)$	$\rho_\mu(y,xy)$	$\rho_\mu(x,xz)$	$\rho_\mu(z,xz)$	$\rho_\mu(y,yz)$	$\rho_\mu(z,yz)$
76/120	44/120	69/120	51/120	55/120	65/120

Table 2: Example 1. [A.] Two type-conditional stochastic choice functions over the binary menus drawn from $X = xyz$. [B.] The mixture choice function derived from these SCFs, assumed to be equally likely, via Equation 4. [C.] The associated marginal SCF.

Example 1. Let $\mathcal{A}$ contain the binary menus drawn from $X = xyz$, and let $\Theta = \{\theta_1, \theta_2\}$ with $\pi(\theta_1) = \pi(\theta_2) = 1/2$. Writing u^θ for the utility function of type θ, assume that this type chooses option w from menu ww' with probability $\rho^\theta(w, ww') = \frac{u^\theta(w)}{u^\theta(w)+u^\theta(w')}$. Setting $u^{\theta_1}(x) = u^{\theta_2}(z) = 3$, $u^{\theta_1}(y) = u^{\theta_2}(x) = 2$, and $u^{\theta_1}(z) = u^{\theta_2}(y) = 1$, we obtain the two type-conditional SCFs displayed in Table 2A. Applying Equation 4 then yields the MCF μ shown in Table 2B; where, for instance, the probability assigned to the choice function $\langle c(xy), c(xz), c(yz)\rangle = \langle x, x, y\rangle$ is calculated as

$$\pi(\theta_1)\rho^{\theta_1}(x,xy)\rho^{\theta_1}(x,xz)\rho^{\theta_1}(y,yz) + \pi(\theta_2)\rho^{\theta_2}(x,xy)\rho^{\theta_2}(x,xz)\rho^{\theta_2}(y,yz)$$
$$= [1/2][3/5][3/4][2/3] + [1/2][2/3][2/5][1/4] = 22/120. \quad (5)$$

The associated marginal SCF, shown in Table 2C, can also be computed as the pointwise average of the two type-conditional SCFs.[11] However, statistical information is lost when we compress ρ^{θ_1} and ρ^{θ_2} into ρ_μ, since the aggregate MCF μ is not menu independent.[12] ‖

The assumption of menu independence conditional on type formalizes the idea that the residual variation captured by the function $\rho^\theta(\cdot, A)$ is uncorrelated across $A \in \mathcal{A}$. In

[11] Here, for example, we have $\rho_\mu(x,xy) = 22/120 + 21/120 + 12/120 + 21/120 = 76/120$; which is the average of $\rho^{\theta_1}(x,xy) = 3/5 = 72/120$ and $\rho^{\theta_2}(x,xy) = 2/3 = 80/120$.

[12] Indeed, μ_{ρ_μ} assigns probability $[76/120][69/120][55/120] \neq 22/120$ to $\langle c(xy), c(xz), c(yz)\rangle = \langle x, x, y\rangle$.

Example 1, the probability that type θ_1 maximizes u^{θ_1} under all three menus is the product $\rho^{\theta_1}(x, xy)\rho^{\theta_1}(x, xz)\rho^{\theta_1}(y, yz) = [3/5][3/4][2/3] = 3/10$ of the maximization probabilities for the individual menus, with "success" in one problem having neither a beneficial nor a detrimental impact on the others. Here we can imagine the agent's level of attention varying due to environmental factors—perhaps hot weather causes discomfort and lack of focus—that are realized independently across menus, though not necessarily with identical distributions. The menus could then represent options in distinct seasons of the year, with systematically different but uncorrelated environmental patterns.

Before proceeding, we comment briefly on the relationship between individuals and population shares in the context of mixture choice data. In our setting, an individual is nothing more than a deterministic choice function $c \in \mathbb{C}$ that encodes his or her behavior across all decision problems in the dataset. Counting the agents that behave in the same way and dividing by the size of the population then yields the corresponding share $\mu(c)$, and conversely multiplying $\mu(c)$ by the population size tells us how many agents behave in the specified manner. We assume no other knowledge of any sort about individuals (in particular, no access to demographics or other decision-maker covariates), and so these two descriptions of the collective behavior are completely equivalent for a population of known size.

In this connection, it is important to note that individuals do not coincide with types. Indeed, in our framework multiple individuals will share the same θ, and types are assumed to be unobservable. If types were observable, then we could pool the agents accordingly to obtain the type distribution and type-conditional SCFs, and it would remain only to determine any deep parameters embedded in each ρ^θ. In contrast, our task in this paper is to use the mixture choice data in μ to identify all components of the model in Equation 4, including both the explicit parameters $\langle \pi(\theta), \rho^\theta \rangle_{\theta \in \Theta}$ and the deep parameters arising from more specialized assumptions about type-conditional choice behavior.

3 Generic identification results: A tool kit

3.1 Sufficient conditions for identification

As noted, we are interested in the identification properties of the general mixture choice model in Equation 4. In particular, our focus is on the number $J = |\Theta|$ of types that the model allows, subject to its parameters being uniquely determined by the observable data. For the moment we address this issue without imposing any additional structure on the model; that is, without making any assumptions about the factors that influence the functions ρ^θ. At present we are content to use our mixture choice data to identify the distributions of types and type-conditional choices, while in Section 4 we will proceed to examine more specialized models for which the identification of "deep" parameters (embedded in ρ^θ) becomes possible. Thus we begin with a tool kit of relatively abstract results, developed in Section 3.

Enumerating the types in Θ as $\langle \theta_1, \ldots, \theta_J \rangle$, the parameters of the model in Equation 4 are $\Omega = \langle \pi(\theta_j), \rho^{\theta_j} \rangle_{j=1}^J$, where $\pi(\theta_j)$ is the probability of type θ_j and ρ^{θ_j} is the associated SCF. The model is (*strictly*) *identified* if the mapping $\Omega \mapsto \mu$ is one-to-one, and *generically identified* if this property holds except on a parameter set with Lebesgue measure zero.[13]

Observe that in Equation 4, the type θ is a dummy variable for the summation. This implies that our enumeration $\langle \theta_1, \ldots, \theta_J \rangle$ of the type space is arbitrary and cannot affect the MCF generated by a given set of parameters—which is to say that any permutation of the type labels leaves μ unchanged. However, as noted by Allman et al. [4, p. 3101], "this does not prevent the statistician from inferring the parameters" of the model, and "identification up to a permutation [of the labels] is largely enough for practical use, at least in a maximum likelihood setting." Our concept of identification must therefore be understood as the map $\Omega \mapsto \mu$ being one-to-one up to reassignment of the type labels.

[13] That is, writing $\triangle^{Z-1}$ for the unit simplex in $\Re^Z$, the exceptional set must have measure zero relative to the parameter space $\triangle^{J-1} \times [\times_{k=1}^K \triangle^{|A_k|-1}]^J \subset \Re^{J+J\sum_{k=1}^K |A_k|}$. (See Lewbel [24] and McManus [33] for discussion of generic identification in econometrics.)

To address the question of identification we shall use sufficient conditions from Allman et al. [4, Theorem 4], based on a fundamental result by Kruskal [23, Theorem 4a]. These conditions may be adapted for our purposes as follows.

Lemma 1 (Allman et al. [4])**.** *Let* $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ *be a partition of* $\mathcal{A}$ *with* $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3 \neq \emptyset$. *Writing* $\kappa_v = \prod_{A \in \mathcal{A}_v} |A|$, *the model is generically identified if*

$$\min\{\kappa_1, J\} + \min\{\kappa_2, J\} + \min\{\kappa_3, J\} \geq 2J + 2. \tag{6}$$

The following example illustrates the use of mixture choice data to identify the type probabilities and type-conditional SCFs.

Example 2. In the setting of Example 1, write $p = \pi(\theta_1)$, $a_j = \rho^{\theta_j}(x, xy)$, $b_j = \rho^{\theta_j}(x, xz)$, and $d_j = \rho^{\theta_j}(y, yz)$. Denoting a given c by the vector $\langle c(xy), c(xz), c(yz)\rangle$, we then have

$$\mu(x, x, y) = pa_1b_1d_1 + [1 - p]a_2b_2d_2, \tag{7}$$

$$\mu(x, x, z) = pa_1b_1[1 - d_1] + [1 - p]a_2b_2[1 - d_2], \tag{8}$$

$$\mu(x, z, y) = pa_1[1 - b_1]d_1 + [1 - p]a_2[1 - b_2]d_2, \tag{9}$$

$$\vdots$$

$$\mu(y, z, z) = p[1 - a_1][1 - b_1][1 - d_1] + [1 - p][1 - a_2][1 - b_2][1 - d_2]. \tag{10}$$

For the partition $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\} = \{\{xy\}, \{xz\}, \{yz\}\}$, we find that Equation 6 holds since $2+2+2 \geq 2\times 2+2$. Accordingly, the multi-linear system in Equations 7–10 is generically solvable for the seven parameters $\langle p, a_1, b_1, d_1, a_2, b_2, d_2\rangle$. For instance, substituting the probabilities from Table 2B into the left-hand-sides of these equations, we obtain a system with solution $\langle p, a_1, b_1, d_1, a_2, b_2, d_2\rangle = \langle 1/2, 3/5, 3/4, 2/3, 2/3, 2/5, 1/4\rangle$, unique up to label reassignment.[14] These are, of course, the parameter values from Table 2A. $\|$

[14]The vector $\langle p', a_1', b_1', d_1', a_2', b_2', d_2'\rangle = \langle 1/2, 2/3, 2/5, 1/4, 3/5, 3/4, 2/3\rangle = \langle 1-p, a_2, b_2, d_2, a_1, b_1, d_1\rangle$ of parameters arises from swapping the two type labels, and hence generates the same MCF.

A second example (for which we are grateful to an anonymous referee) illustrates parameters in the zero-measure set on which identification fails.[15]

Example 3. Let $\mathcal{A} = \{xyz, xy, yz\}$ and $\Theta = \{\theta_1, \theta_2\}$, and denote a given c by the vector $\langle c(xyz), c(xy), c(yz)\rangle$. Write $p = \pi(\theta_1)$, $a_j = \rho^{\theta_j}(x, xyz)$, $b_j = \rho^{\theta_j}(y, xyz)$, $d_j = \rho^{\theta_j}(x, xy)$, and $e_j = \rho^{\theta_j}(y, yz)$. The parameter vectors

$$\langle p, a_1, b_1, d_1, e_1, a_2, b_2, d_2, e_2\rangle = \langle 3/5, 1/2, 1/2, 1, 1, 1, 0, 1, 7/10\rangle, \tag{11}$$

$$\langle p', a_1', b_1', d_1', e_1', a_2', b_2', d_2', e_2'\rangle = \langle 7/10, 1, 0, 1, 29/35, 0, 1, 1, 1\rangle; \tag{12}$$

then both generate the MCF with $\mu(xxy) = 29/50$, $\mu(xxz) = 6/50$, $\mu(yxy) = 15/50$. $\|$

3.2 How many types can be identified?

We now use Lemma 1 to investigate the question of how many types could conceivably be distinguished by mixture choice data from our collection $\mathcal{A}$ of menus. Denote by $\overline{J}(\mathcal{A})$ the largest number of types for which the lemma guarantees identification. This value is obtained by choosing the partition $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ optimally, so as to maximize the largest value of J that satisfies the inequality in Equation 6.

Throughout Section 3, we assume without loss of generality that $|A_1| \geq \cdots \geq |A_K|$. Clearly $|A_1| \leq n$ and $|A_K| \geq 2$, but neither constraint necessarily holds with equality for arbitrary $\mathcal{A}$. In Equation 6, any partition that achieves $\overline{J}(\mathcal{A})$ will assign a menu with the smallest available cardinality to its own element, which we express as $\mathcal{A}_3 = \{A_K\}$. Hence $\kappa_3 = |A_K| \geq 2$, the desired inequality is implied by $\min\{\kappa_1, \kappa_2\} \geq J$, and we can write

$$\overline{J}(\mathcal{A}) = \max_{\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{A} \setminus \mathcal{A}_3} \min\{\kappa_1, \kappa_2\} \tag{13}$$

for the resulting upper bound.

[15] Indeed, this example can be shown to fail the rank condition in Kruskal [23, Theorem 4a]. Note that the lack of identification here is not due to the zero probabilities in Equations 11–12. Similar examples can be adduced in which the type-conditional SCFs return only strictly positive values.

Ideally we would partition $\mathcal{A}\backslash\mathcal{A}_3$ so as to set $\kappa_1 = \kappa_2$, enabling the upper bound $\overline{J}(\mathcal{A})$ in Equation 13 to attain its "theoretical optimum" value

$$J^*(\mathcal{A}) = \left[\textstyle\prod_{k=1}^{K-1} |A_k|\right]^{1/2}. \tag{14}$$

However, achieving this optimum requires finding a set $\Xi \subset \{1, \ldots, K-1\}$ of menu indices such that $\prod_{k\in\Xi} |A_k| = J^*(\mathcal{A})$, which may or may not be possible for any particular $\mathcal{A}$. Thus it remains challenging in general to determine the value of the bound $\overline{J}(\mathcal{A})$.[16]

3.3 All menus

As a benchmark, we first consider the prospects for identification in the ideal case where our dataset includes the full collection of menus. For instance, when $n = 4$ we have that $K = 2^n - n - 1 = 11$ and the list of menu cardinalities is $\langle 4, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2\rangle$. Here setting $\mathcal{A}_1 = \{A_1, A_2, A_6, A_7, A_8\}$ yields $\kappa_1 = 4\cdot3\cdot2\cdot2\cdot2 = 96$ and $\kappa_2 = 3\cdot3\cdot3\cdot2\cdot2 = 108$; whereupon $\min\{\kappa_1, \kappa_2\} = 96$ and it is easy to check that no higher value is achievable. We record this conclusion as follows.

Proposition 1. *If $n = 4$, $\mathcal{A}$ is the full collection of menus, and $J \leq 96$, then the model is generically identified.*

In general there are $\binom{n}{m}$ menus of each size m, with one binary menu allocated to $\mathcal{A}_3$, and so the theoretical optimum $J^*(\mathcal{A})$ satisfies

$$\log J^*(\mathcal{A}) = \frac{1}{2}\log\left[\frac{1}{2}\prod_{m=2}^{n} m^{\binom{n}{m}}\right] = \frac{1}{2}\left[\log\frac{1}{2} + \sum_{m=2}^{n}\binom{n}{m}\log m\right]. \tag{15}$$

When $n \geq 5$, we can use an adaptive algorithm to identify a pair of sets $\mathcal{A}_1$ and $\mathcal{A}_2$ that may not achieve the true $\overline{J}(\mathcal{A})$, but are certain to yield a value of $\min\{\kappa_1, \kappa_2\}$ that is at

[16] A closely related problem in computer science is that of separating a given multiset of numbers into two subsets in such a way as to minimize the difference between the sums of the numbers in each subset. This "partition problem" is known to be NP-hard.

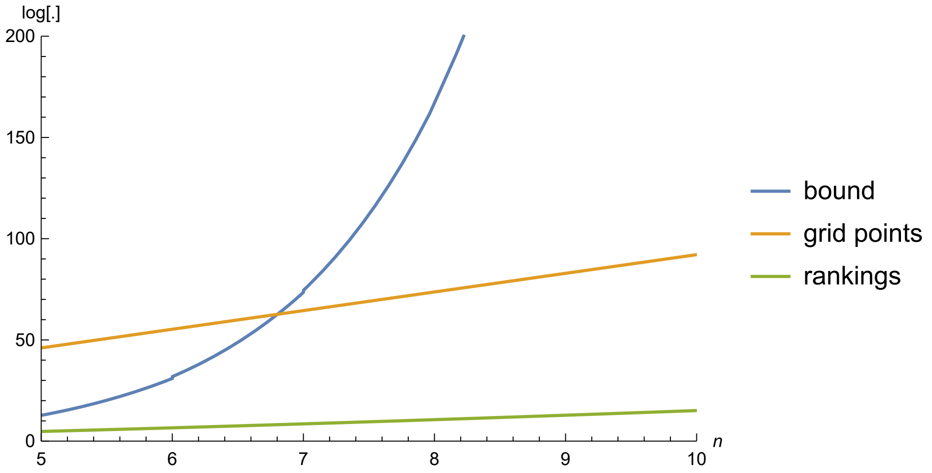

Figure 1: For $5 \leq n \leq 10$, the bound $\left[\frac{n-2}{n-1}\right]^{1/2} J^*(\mathcal{A})$ from Proposition 2, the number 10^{4n} of points on a cardinal preference grid with four significant figures, and the number $n!$ of ordinal preference rankings (on a logarithmic scale).

least the "near-optimum" $\left[\frac{n-2}{n-1}\right]^{1/2} J^*(\mathcal{A})$. The following result (proved in Appendix A.1) establishes this claim.

Proposition 2. *If $n \geq 5$, $\mathcal{A}$ is the full collection of menus, and*

$$\log J \leq \frac{1}{2} \log \frac{n-2}{n-1} + \log J^*(\mathcal{A}) = \frac{1}{2}\left[\log \frac{n-2}{2[n-1]} + \sum_{m=2}^{n} \binom{n}{m} \log m\right], \tag{16}$$

then the model is generically identified.

Proposition 2 establishes that models with a large number of types are in principle identifiable given mixture choice data from all menus. For $5 \leq n \leq 10$, Figure 1 plots the bound on J supplied by Equation 16. Also shown are the number 10^{4n} of points on the grid that results from each of n alternatives being assigned a cardinal utility value with four significant figures, as well as the number $n!$ of ordinal (strict) preference rankings of n options. As the figure illustrates, by $n = 7$ the bound is several orders of magnitude above the cardinality of either of these generously proportioned preference-type spaces, and by $n = 8$ it is vastly higher. Thus the message of Proposition 2 is that with mixture choice data from all menus over even a modest number of alternatives, (generic) identification is unlikely to be an issue for the model in Equation 4. The more relevant question will be when theoretical identification can be ensured for rich type spaces despite limitations on the available data.

As applied to choice experiments, Proposition 2 underscores the importance of taking identification into account at the design stage. The menus from which subjects choose must of course be selected in advance, but the grids used to approximate any continuous parameters may sometimes be selected later, when estimation is performed.[17] Our result links these two features, informing us which menu collections are in principle adequate to identify numerical parameters with specified precision.

3.4 Menus of uniform size

Next we consider restricted datasets containing at least K_q menus of uniform size q. Since $K_q - 1$ of these menus are available to be divided between $\mathcal{A}_1$ and $\mathcal{A}_2$, we have that $\kappa_1\kappa_2 \geq q^{K_q-1}$ and $\min\{\kappa_1, \kappa_2\} \geq q^{\lfloor [K_q-1]/2 \rfloor}$ (where $\lfloor \cdot \rfloor$ denotes the "floor" operator). This yields sufficient conditions similar to those in the Allman et al. [4, pp. 3110–3111] application of Lemma 1 to finite mixtures of finite measure products.

Proposition 3. *For $q \leq n-1$, let $\mathcal{A}$ contain at least $K_q \leq \binom{n}{q}$ menus of cardinality q. Then the model is generically identified if* $\log J \leq \lfloor \frac{1}{2}[K_q - 1] \rfloor \log q$.

Corollary 1. *For $q \leq n-1$, let $\mathcal{A}$ contain all menus of cardinality q. Then the model is generically identified if* $\log J \leq \lfloor \frac{1}{2}[\binom{n}{q} - 1] \rfloor \log q$.

For instance, the case $q = n-1$ corresponds to "leave-one-out" menu variation of the sort used by Abaluck and Adams [1] to identify parameters in an econometric specification of limited attention models. With $K_q = \binom{n}{n-1}$, so that all leave-one-out menus are in $\mathcal{A}$, the bound in Corollary 1 is $\lfloor \frac{1}{2}[n-1] \rfloor \log[n-1]$. Here mixture choice data will allow us to identify 16 types with 5 options, 216 with 7 options, and 4096 with 9 options; indicating the power of this sort of variation relative to the all-menus benchmark.

[17] We thank an anonymous referee for making this point.

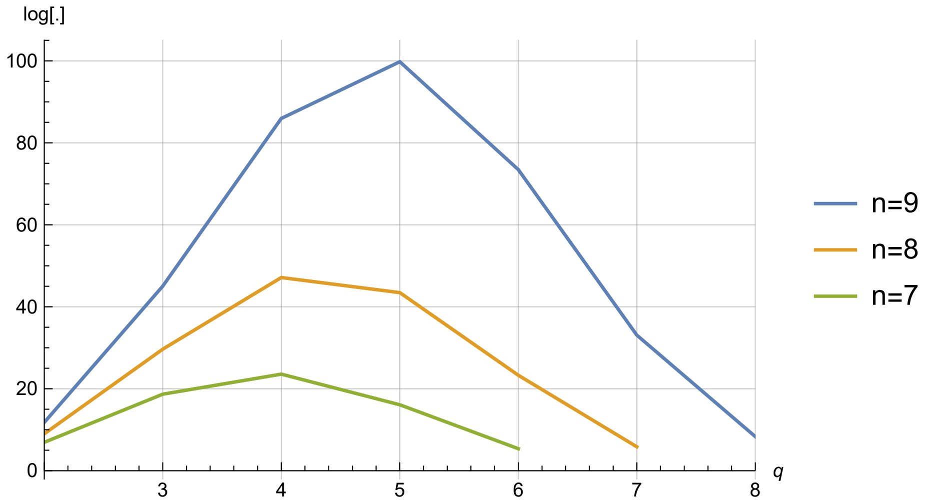

Figure 2: For $7 \leq n \leq 9$ and $q < n$, the bound $\lfloor \frac{1}{2}[\binom{n}{q} - 1] \rfloor \log q$ from Corollary 1.

On the other hand, for either theoretical or practical reasons a researcher may instead wish to work with low-cardinality menus. This might be necessary, for example, if data comes in the form of putative preference comparisons—equivalent to binary choices—or if there is a desire to avoid context or choice-overload effects. Setting $q = 2$ and $K_q = \binom{n}{2}$, we obtain the bound $\lfloor \frac{1}{4}[n-2][n+1] \rfloor \log 2$; thus identifying 16 types with 5 options, 1024 with 7 options, and 131 072 with 9 options.

Observe that the set of all binary menus and the set of all leave-one-out menus both use a total of $n[n-1]$ choice objects across all problems. More generally, the set of all menus of a given size $q \leq n/2$ and the set of all menus of size $n - q + 1$ both use a total of $q\binom{n}{q}$ choice objects across all problems. Indeed, each specific alternative appears the same number of times in the two datasets, which are thus directly comparable in terms of resource intensity. In all cases the dataset containing the smaller menus (of size q) allows identification of more preference types than does the dataset containing the larger menus (of size $n - q + 1$). This is illustrated in Figure 2, where for n between 7 and 9 and each $q \leq n$ we plot (on a logarithmic scale) the bound on J from Corollary 1. For example, when $n = 9$ the bound for size $q = 4$ exceeds that for size $n - q + 1 = 6$, when $n = 8$ the bound for $q = 3$ exceeds that for $n - q + 1 = 6$, and so on. Intuitively, this is because a fixed budget B of choice objects is best spent "purchasing" a large number B/q of menus with small size q in order to maximize the bound $\lfloor \frac{1}{2}[\frac{B}{q} - 1] \rfloor \log q$ in Proposition 3.[18]

As already noted, an important threshold for type identification is the number $n!$ of ordinal preference rankings. Given q, we can ask whether the set of all menus of this size is adequate to attain the ranking threshold, and can answer in the affirmative if n is large enough to satisfy the inequality $\log n! \leq \lfloor \frac{1}{2}[\binom{n}{q} - 1] \rfloor \log q$ arising from Corollary 1. The next result records our conclusions from this exercise.

Proposition 4. *Suppose that $J = n!$ and $\mathcal{A}$ consists of all menus of cardinality q. Then the model is generically identified if and only if one of the following holds:* **(i)** *$q = 2$ and $n \geq 10$,* **(ii)** *$q = 3$ and $n \geq 6$, or* **(iii)** *$q \geq 4$ and $n \geq q + 2$.*

Identification of all ordinal preference types is achieved with the smallest number $K = 15$ of menus in the case of $n = 6$ and $q = 4$; and is achieved with the smallest budget $B = 60$ of choice objects in the case of $n = 6$ and $3 \leq q \leq 4$. Of course, if we combine the menus of two or more different cardinalities (e.g., $q = 2$ or 3; see Section 4.4), then identification will be achieved all the more easily.

3.5 Arbitrary menus

With naturalistic datasets having no special structure, we can still obtain a useful bound on the number of identifiable types, relative to the theoretical optimum $J^*(\mathcal{A})$, as long as the cardinalities of the menus in $\mathcal{A}$ do not decline too quickly.

Definition 2. The collection $\mathcal{A}$ is *subquadratic* if for $k < K - 1$ we have $|A_k| \leq |A_{k+1}|^2$.

Proposition 5. *If $\mathcal{A}$ is subquadratic and* $\log J \leq \frac{1}{2} \log \frac{1}{|A_{K-1}|} + \log J^*(\mathcal{A})$*, then the model is generically identified.*

[18]Ignoring both the floor operator and integer constraints, we have $\frac{\partial}{\partial q}[\frac{1}{2}[\frac{B}{q} - 1] \log q] = \frac{B[1 - \log q]}{2q^2} - \frac{1}{2q}$, which is strictly negative when $\log q > 1$. As the budget B becomes large, the ideal menu size converges to $e \approx 2.7$ from below.

This result shows, for instance, that if $\mathcal{A}$ is subquadratic and the two smallest menus in the collection are binary, then we know that at least a fraction $2^{-1/2} \approx 0.707$ of $J^*(\mathcal{A})$ is attainable as a bound. (Proposition 5 is proved in Appendix A.1.)

4 Specialized models of choice

4.1 Deep parameters

The results in Section 3 ensure that the parameters $\Omega = \langle \pi(\theta_j), \rho^{\theta_j} \rangle_{j=1}^{J}$ of the model in Equation 4 can be generically identified from mixture choice datasets. In more specialized applications of this methodology, however, the ultimate objects of interest will often be the "deep" parameters that impact decision making conditional on the agent's type. Our framework embeds these parameters in the functions ρ^{θ_j}, from which we must be able to deduce them if the specialized model is to be identified.

In this section we interpret types as ordinal (strict) preference rankings, and for this reason $J = |\Theta| = n!$ will hold throughout. Here we examine the issue of deep parameter identification in relation to two specific models of cognitively constrained decision making; namely, random satisficing thresholds and "quantal Fechnerian" choice. In the first model the agent does not always aspire to perfect utility maximization, while in the second the frequencies of choice errors depend on cardinal utility differences. Although many other specialized models could be treated in the same way, we limit attention to these two for the sake of brevity.[19]

As noted, demonstrating that a specialized model is identified will require exhibiting a collection of menus sufficient not only to reveal the type-conditional SCFs, but also to elicit the relevant deep parameters. Furthermore, we must establish that this can be done generically, not in the space of all models consistent with Equation 4, but rather in

[19] Deep parameter identification would of course continue to be relevant outside the context of preference order types and boundedly rational choice conditional on preferences. For instance, we could have a model with types as frames and type-dependent risk tolerance parameters to be identified.

the lower-dimensional space of models consistent with our assumptions about cognition. Put differently, we must confirm that the specialized model does not itself fall into the non-identifiable gaps allowed by Lemma 1.

In Sections 4.2–4.3 we briefly define the two specialized models under consideration. Section 4.4 then proceeds to state a generic identification result for these models, showing that both preferences and cognitive parameters are revealed by mixture choice data from the collection of binary and ternary menus.

4.2 Random satisficing thresholds

As first conceived by Herbert Simon, a "satisficer" is an agent who chooses an alternative that is not necessarily optimal, but exceeds some threshold level of utility that is deemed acceptable. To capture this idea, let type θ have preference order $\succ^\theta$ and write $\succeq^\theta$ for its union with the equality relation. Each available $\tilde{x}$ may potentially be realized as the threshold alternative for menu A, and this occurs with a probability $\tau_A^\theta(\tilde{x}) > 0$ that in general depends on both the menu and the agent's type. Any options dispreferred to the threshold will not be selected, and for simplicity we assume that all other alternatives are equally likely to be the final choice. Denoting by $r_A^\theta(\tilde{x}) = \left|\left\{y \in A : y \succeq^\theta \tilde{x}\right\}\right|$ the rank of alternative $\tilde{x}$ on menu A according to the preference order $\succ^\theta$, this yields the conditional SCF given by

$$\rho^\theta(x, A) = \sum_{\tilde{x} \in A:\, x \succeq^\theta \tilde{x}} \frac{\tau_A^\theta(\tilde{x})}{r_A^\theta(\tilde{x})} \tag{17}$$

(cf. Dardanoni et al. [14, Appendix B, pp. 3–4]). Here the numerator of the summand is the probability that a particular $\tilde{x}$ is the realized threshold, while the denominator is the number of satisfactory alternatives (relative to the current threshold) between which this probability is equitably shared.

4.3 Quantal Fechnerian choice

Let $\mathcal{A}$ contain all possible binary menus. Strzalecki [37] defines a stochastic choice function ρ to be *Fechnerian* if it admits a utility function $u : X \to \Re$ and a strictly increasing map φ such that for all $x, y \in X$ we have $\rho(x, xy) = \varphi(u(x) - u(y))$.[20] That is to say, binary choice probabilities are determined by utility differences.[21]

We will use an ordinal analog of the Fechnerian class of SCFs in which, for "decay" parameter $\delta_A^\theta \in (0, 1)$, the probability of alternative x being chosen from menu A is

$$\rho^\theta (x, A) = [\delta_A^\theta]^{r_A^\theta(x)-1} \times \frac{1 - \delta_A^\theta}{1 - [\delta_A^\theta]^{|A|}}. \tag{18}$$

Hence $\rho^\theta (x, A) = [\delta_A^\theta]^{r_A^\theta(x) - r_A^\theta(y)} \times \rho^\theta (y, A)$, and the probability of an option being chosen declines geometrically with its preference rank. Because of this dependence on ordinal rather than cardinal preference information, and because of the resulting discreteness of changes in choice probability, Equation 18 will be referred to as the "quantal" Fechnerian model.

4.4 Identification of specialized models

Our aim is to provide sufficient conditions for generic identification of deep parameters under either random satisficing thresholds (RST) or quantal Fechnerian choice (QFC). In the RST context, this will mean two things: First, the parameters $\langle \succ^\theta, \langle \tau_A^\theta(\tilde{x}) \rangle_{\tilde{x} \in A \in \mathcal{A}} \rangle$ of Equation 17 must be revealed by the stochastic choice data in ρ^θ. Second, the parameters $\Omega = \langle \pi(\theta_j), \rho^{\theta_j} \rangle_{j=1}^J$ of Equation 4 must be identified generically on the class of MCFs for which each ρ^θ is consistent with the RST assumptions. Similarly, in the QFC context we need that the parameters $\langle \succ^\theta, \langle \delta_A^\theta \rangle_{A \in \mathcal{A}} \rangle$ of Equation 18 are revealed by ρ^θ and that those

[20] Gustav Fechner [16], the originator of "psychophysics," investigated the relationship between intensity and perception of stimuli. Fechnerian models of choice were later studied by Thurstone [38].

[21] For instance, the standard multinomial logit model is Fechnerian, with $\varphi(v) = [1 + \exp[-v]]^{-1}$. This property is also satisfied by a special case of the random threshold model, proposed by Tyson [40], with $\varphi(v) = 1 - \frac{1}{2}\exp[-v]$ for $v \geq 0$ and $\varphi(v) = \frac{1}{2}\exp v$ for $v < 0$.

of Equation 4 are identified generically on the corresponding class of MCFs.[22]

We address both the RST and QFC models in a single result, using a common menu collection.[23]

Proposition 6. *Let $n \geq 4$, and let $\mathcal{A}$ contain all possible binary and ternary menus. In the RST model, the type-conditional SCFs ρ^θ generically identify the preference orders $\succ^\theta$ and threshold probabilities $\tau_A^\theta(\tilde{x})$. Likewise, in the QFC model the functions ρ^θ generically identify the preference orders $\succ^\theta$ and decay parameters δ_A^θ.*

The proof of Proposition 6 (supplied in Appendix A.2) follows that of Lemma 1 by Allman et al. [4], although the two results are formally independent. After showing that the deep parameters of each model are revealed by the type-conditional SCFs, we proceed to separate $\mathcal{A}$ into all menus of cardinality three (partition element $\mathcal{A}_1$), a single menu of cardinality two ($\mathcal{A}_3$), and the remaining menus ($\mathcal{A}_2$). For each partition element $\mathcal{A}_v$ and for given threshold distributions over the menus therein, we then construct a matrix Φ^v whose entries are mixture choice probabilities for the possible deterministic choice functions over $\mathcal{A}_v$ (rows of the matrix) and the possible preference types (columns of the matrix) under the RST model. The desired identification will then hold if these matrices generically have full "Kruskal rank"; or, equivalently, if they generically possess a nonzero minor determinant of order J (the number of columns).[24] For the RST model, the minor determinants of the Φ^v matrices are polynomials in the threshold parameters. Using the fact that a polynomial function on a Euclidean space is either identically zero or nonzero

[22] Recall that the parameter space of the general mixture choice model is $\triangle^{J-1} \times [\times_{A\in\mathcal{A}} \triangle^{|A|-1}]^J$ (see fn. 13). The more specialized RST model can be viewed as an algebraic variety in this space, consisting of those vectors $\langle \pi(\theta_j), \rho^{\theta_j} \rangle_{j=1}^J$ for which there exist deep parameters $\langle \tau_A^{\theta_j}(\tilde{x}) \rangle_{\tilde{x}\in A\in\mathcal{A}}$ that produce each ρ^{θ_j} via Equation 17. This set is a proper subvariety that is spanned by the parameters $\tau_A^{\theta_j}(\tilde{x})$ of the threshold distributions—together with the parameters $\pi(\theta_j)$ of the type distribution—and we wish to establish that identification holds generically on this structured class of MCFs. (Similar statements hold for the QFC model, the deep parameters $\langle \delta_A^{\theta_j} \rangle_{A\in\mathcal{A}}$, and Equation 18.)

[23] Proposition 6 achieves identification using a domain of small (binary and ternary) menus, but other domains might be called for depending on the research question. For instance, a study of choice overload (such as the famous jam-merchandising experiment by Iyengar and Lepper [20]) might require a domain including both small and large menus, and Proposition 6 would need to be modified accordingly.

[24] The Kruskal rank of a matrix is the largest number ρ such that any ρ columns drawn from the matrix are linearly independent. This value is clearly less than or equal to the (ordinary) rank of the matrix.

almost everywhere, we can ensure generic identification by exhibiting specific parameters that make a specific determinant (of order J) nonzero. Moreover, with slight modifications this argument goes through for the QFC model as well.

It is worth emphasizing that Lemma 1 cannot be used directly to prove Proposition 6, due to the reduction in free parameters and the special functional forms of the RST and QFC models. While the broad proof strategy and the definition of the Φ^v matrices follow Allman et al. [4], showing the (generic) full Kruskal rank property in the context of these specialized models requires new arguments and constructions. In particular, our Lemma 2 (in Appendix A.2) plays the role of Allman et al.'s [4, p. 3119] Lemma 13, ensuring full Kruskal rank by exhibiting the parameters and associated nonzero determinants called for above. This argument can potentially be applied to other similar cases, and perhaps even generalized to an overarching class of models for which our approach is valid—a research program that goes beyond the scope of the present paper.

5 Conclusion

In this paper we use mixture choice functions to represent a rich form of data that can arise in both experimental and non-experimental contexts. The essential feature of this data is that it records the *joint* distribution of choices of a population of agents across a series of decision problems, or "occasions" (see Dardanoni et al. [14, p. 1285]). A novelty of our formulation is that we structure the occasions as distinct menus, comprising the domain of a deterministic choice function, which leads to the MCF primitive in Definition 1.

Our formal results concern the identification properties of the model in Equation 4. First, we explore the implications for our framework of the sufficient conditions supplied by Allman et al. [4] (reproduced in Lemma 1 above), establishing in Propositions 1–5 that various numbers and sizes of menus combine to ensure generic identification of both the type distribution and the type-conditional SCFs. In Proposition 6 we then proceed to show that this methodology can be adapted to deal with specialized, lower-dimensional models in which the type-conditional SCFs have particular functional forms and involve deep parameters that must themselves be identified. (Identification having been verified, Appendix A.3 sketches how structural parameters would be estimated from our mixture choice dataset.)

We emphasize that the results in Sections 3 and 4 relate to Lemma 1 in different ways. In Section 3 we are concerned with whether the menu collection is rich enough to identify the parameters of our general framework, and here we can use Lemma 1 directly. In Section 4, on the other hand, our goal is to adapt Allman et al.'s proof so that it can be applied to the specialized models of interest. Both types of results will be needed if we are to work with a variety of datasets and models, and hence these two parts of the paper are complementary.

As noted, we conceive of the choice occasions as distinct menus of alternatives. Other interpretations are possible, however, and will lead to datasets that can be analyzed with similar methods. For example, we could interpret the occasions as repeated choices from a menu that does not change in any consequential way, but whose framing (e.g., merchandising presentation or list order) varies over time. This would give the dataset a behavioral flavor and offer the prospect of "market research" applications, such as using retail scanner data to infer the characteristics of a population of consumers in terms of both preferences and susceptibility to manipulation. Here covariates may be absent (e.g., due to privacy concerns) or unreliable (e.g., due to elicitation by survey), but the link between choices on different occasions will often be observed directly.[25] Studying such variations on the MCF concept is left for future work.

[25]In relation to mixture choice data and privacy concerns, see for instance Anderson et al. [5, p. 421]. Describing the empirical basis for their study of repeated consumption in online settings, these authors explain that a typical entry in one of their datasets consists of "a single user's complete consumption history in chronological order[.]" They go on to write that "[s]ince our focus is on aggregate behavior, no user identities are present in our data, and precautions were taken so that they cannot be recovered[.]"

A Appendix

A.1 Proofs of Propositions 2 and 5

In this section we prove Propositions 2 and 5, stated in Section 3.[26] To do so, we employ a "greedy" partition of $\mathcal{A}$ that sets $\mathcal{A}_3 = \{A_K\}$ and assigns the remaining menus to $\mathcal{A}_1$ and $\mathcal{A}_2$ recursively, according to which of these partition elements has the smaller product-of-cardinalities at the point of assignment. Recalling that $|A_1| \geq \cdots \geq |A_K|$, this partition can be defined formally as follows.

Definition 3. Let $\mathcal{B}_1(1) = \{A_1\}$ and $\mathcal{B}_2(1) = \emptyset$, and for $1 \leq k \leq K-2$ define the sets

$$\mathcal{B}_1(k+1) = \begin{cases} \mathcal{B}_1(k) \cup \{A_{k+1}\} & \text{if } \prod_{B_1 \in \mathcal{B}_1(k)} |B_1| \leq \prod_{B_2 \in \mathcal{B}_2(k)} |B_2|, \\ \mathcal{B}_1(k) & \text{otherwise;} \end{cases} \tag{19}$$

$$\mathcal{B}_2(k+1) = \{A_1, \ldots, A_{k+1}\} \setminus \mathcal{B}_1(k+1). \tag{20}$$

The *greedy partition* of $\mathcal{A}$ then has $\mathcal{A}_1 = \mathcal{B}_1(K-1)$, $\mathcal{A}_2 = \mathcal{B}_2(K-1)$, and $\mathcal{A}_3 = \{A_K\}$.

For $1 \leq k \leq K-1$, let $b_1(k) = \prod_{B_1 \in \mathcal{B}_1(k)} |B_1|$ and $b_2(k) = \prod_{B_2 \in \mathcal{B}_2(k)} |B_2|$, and define the product-of-cardinalities ratio

$$R(k) = \frac{\max\{b_1(k), b_2(k)\}}{\min\{b_1(k), b_2(k)\}} \geq 1. \tag{21}$$

To prove Proposition 5, we demonstrate that $\min\{\kappa_1, \kappa_2\} \geq |A_{K-1}|^{-1/2} J^*(\mathcal{A})$ under the greedy partition of $\mathcal{A}$.

Proof of Proposition 5. We first show by induction that each $R(k) \leq |A_k|$. Assuming that this inequality holds for a particular k, consider the ratio $R(k+1)$. On the one hand, if $|A_{k+1}| \geq R(k)$ then $R(k+1) = |A_{k+1}| / R(k) \leq |A_{k+1}|$, since $R(k) \geq 1$ by construction.

[26] Since Proposition 5 imposes weaker assumptions on the menu collection $\mathcal{A}$, we prove this result first.

On the other hand, if $|A_{k+1}| < R(k) \leq |A_k|$ then

$$R(k+1) = \frac{R(k)}{|A_{k+1}|} \leq \frac{|A_k|}{|A_{k+1}|} \leq |A_{k+1}|\,, \tag{22}$$

where the last inequality holds because $\mathcal{A}$ is subquadratic. To begin the inductive chain, observe that $R\,(1) = |A_1|\,/1 = |A_1|$.

By induction we have $R(K-1) \leq |A_{K-1}|$, and it follows that

$$|A_{K-1}| \cdot \min\{b_1(K-1), b_2(K-1)\} \geq \max\{b_1(K-1), b_2(K-1)\}\,, \tag{23}$$

$$|A_{K-1}| \cdot \min\{\kappa_1, \kappa_2\} \geq \max\{\kappa_1, \kappa_2\}\,, \tag{24}$$

$$|A_{K-1}| \cdot [\min\{\kappa_1, \kappa_2\}]^{\,2} \geq \kappa_1\kappa_2 = [J^*(\mathcal{A})]^{\,2}, \tag{25}$$

$$\min\{\kappa_1, \kappa_2\} \geq |A_{K-1}|^{-1/2}\, J^*(\mathcal{A}), \tag{26}$$

as desired. □

To establish Proposition 2, we start with the greedy partition and then judiciously exchange menus between $\mathcal{A}_1$ and $\mathcal{A}_2$ to force their product-of-cardinalities ratio as close as possible to unity.

Proof of Proposition 2. Observe first that under the greedy partition, $\mathcal{A}_1$ and $\mathcal{A}_2$ each contain at least two menus with every cardinality from 2 to $n-1$. This is because when the cardinality decreases from $|A_k|$ to $|A_{k+1}| = |A_k| - 1 \geq 2$, at most two consecutive menus can be allocated to the same partition element before the assignment process begins to alternate between $\mathcal{A}_1$ and $\mathcal{A}_2$ (for the remainder of the block of menus with cardinality $|A_{k+1}|$). Indeed, once the first two menus with cardinality $|A_{k+1}|$ have been allocated to the same partition element, we must have $R(k) < |A_{k+1}|^2$; since otherwise

$$1 \leq R(k+2) = \frac{R(k)}{|A_{k+1}|^2} \leq \frac{|A_k|}{|A_{k+1}|^2} = \frac{|A_{k+1}|+1}{|A_{k+1}|^2} \leq \frac{3}{4}, \tag{27}$$

a contradiction. Noting that $\binom{n}{m} \geq n \geq 5$ for $2 \leq m \leq n-1$, the observation follows.

Recall that the greedy partition has $\max\{\kappa_1, \kappa_2\}/\min\{\kappa_1, \kappa_2\} \leq 2$, by Equation 24. If also $\max\{\kappa_1, \kappa_2\}/\min\{\kappa_1, \kappa_2\} > [n-1]/[n-2]$, then let m be the unique integer such that both $3 \leq m \leq n-1$ and

$$\frac{m}{m-1} < \frac{\max\{\kappa_1, \kappa_2\}}{\min\{\kappa_1, \kappa_2\}} \leq \frac{m-1}{m-2}. \tag{28}$$

Now choose menus $B_1 \in \mathcal{A}_1$ and $B_2 \in \mathcal{A}_2$ such that $|B_1| = m$ and $|B_2| = m-1$ if $\kappa_1 > \kappa_2$, whereas $|B_1| = m-1$ and $|B_2| = m$ if $\kappa_1 < \kappa_2$. Using the selected menus, we define a new partition with elements $\mathcal{A}'_1 = [\mathcal{A}_1 \setminus \{B_1\}] \cup \{B_2\}$, $\mathcal{A}'_2 = [\mathcal{A}_2 \setminus \{B_2\}] \cup \{B_1\}$, and $\mathcal{A}'_3 = \mathcal{A}_3$. If $\kappa_1 > \kappa_2$, then the new κ'_1 and κ'_2 satisfy

$$\frac{m-1}{m} < \frac{\kappa'_1}{\kappa'_2} = \frac{\kappa_1}{\kappa_2}\left[\frac{m-1}{m}\right]^2 \leq \frac{[m-1]^3}{m^2[m-2]} < \frac{m}{m-1}, \tag{29}$$

where the equality holds by construction, the first and second inequalities follow from Equation 28, and the third inequality holds since $m \geq 3$. The case of $\kappa_1 < \kappa_2$ is analogous, and in either eventuality we have that $\max\{\kappa'_1, \kappa'_2\}/\min\{\kappa'_1, \kappa'_2\} < m/[m-1]$. Once again, if $\max\{\kappa'_1, \kappa'_2\}/\min\{\kappa'_1, \kappa'_2\} > [n-1]/[n-2]$ then there is a new integer m' such that both $m+1 \leq m' \leq n-1$ and

$$\frac{m'}{m'-1} < \frac{\max\{\kappa'_1, \kappa'_2\}}{\min\{\kappa'_1, \kappa'_2\}} \leq \frac{m'-1}{m'-2}. \tag{30}$$

Choosing menus $B'_1 \in \mathcal{A}'_1$ and $B'_2 \in \mathcal{A}'_2$ such that $|B'_1| = m'$ and $|B'_2| = m'-1$ if $\kappa'_1 > \kappa'_2$, whereas $|B'_1| = m'-1$ and $|B'_2| = m'$ if $\kappa'_1 < \kappa'_2$, we can proceed as before to construct a third partition. Moreover, at most two menus with any particular cardinality need to be found in $\mathcal{A}_1$ and $\mathcal{A}_2$ (for instance, if $|B_1| = m$, $m' = m+1$, and $|B'_1| = m'-1 = m$).

Repeating the above procedure a finite number of (at most $n-3$) times, we obtain a partition with elements $\mathcal{A}_1^{\prime\cdots\prime}$, $\mathcal{A}_2^{\prime\cdots\prime}$, and $\mathcal{A}_3^{\prime\cdots\prime}$; and with corresponding $\kappa_1^{\prime\cdots\prime}$ and $\kappa_2^{\prime\cdots\prime}$ that satisfy $\max\{\kappa_1^{\prime\cdots\prime}, \kappa_2^{\prime\cdots\prime}\}/\min\{\kappa_1^{\prime\cdots\prime}, \kappa_2^{\prime\cdots\prime}\} \leq [n-1]/[n-2]$. We then have

$$[n-1]\min\{\kappa_1^{\prime\cdots\prime}, \kappa_2^{\prime\cdots\prime}\} \geq [n-2]\max\{\kappa_1^{\prime\cdots\prime}, \kappa_2^{\prime\cdots\prime}\}, \tag{31}$$

$$[n-1][\min\{\kappa_1^{\prime\cdots\prime}, \kappa_2^{\prime\cdots\prime}\}]^2 \geq [n-2]\kappa_1^{\prime\cdots\prime}\kappa_2^{\prime\cdots\prime} = [n-2][J^*(\mathcal{A})]^2, \tag{32}$$

$$\min\{\kappa_1^{\prime\cdots\prime}, \kappa_2^{\prime\cdots\prime}\} \geq \left[\tfrac{n-2}{n-1}\right]^{1/2} J^*(\mathcal{A}); \tag{33}$$

and the result follows. □

A.2 Proof of Proposition 6

In this section we prove Proposition 6, stated in Section 4.

Proof of Proposition 6. We demonstrate first that the deep parameters of each specialized model are revealed by a given collection of type-conditional SCFs.

For the RST model, $x \succ^\theta y$ implies $\rho^\theta(x, xy) - \rho^\theta(y, xy) = \tau^\theta_{xy}(x) > 0$, and conversely $\rho^\theta(x, xy) - \rho^\theta(y, xy) > 0$ only if $x \succ^\theta y$. Moreover, for $x, y, z \in A$ with $r^\theta_A(y) = r^\theta_A(x) + 1$ and $r^\theta_A(z) = |A|$ we have $\tau^\theta_A(x) = r^\theta_A(x)[\rho^\theta(x, A) - \rho^\theta(y, A)]$ and $\tau^\theta_A(z) = |A| \times \rho^\theta(z, A)$. Thus $\succ^\theta$ and each τ^θ_A are revealed by ρ^θ.

For the QFC model, $x \succ^\theta y$ implies $\rho^\theta(x, xy)/\rho^\theta(y, xy) = [\delta^\theta_A]^{-1} > 1$, and conversely $\rho^\theta(x, xy)/\rho^\theta(y, xy) > 1$ only if $x \succ^\theta y$. Moreover, for $x, y \in A$ with $r^\theta_A(y) = 2 > 1 = r^\theta_A(x)$ we have $\delta^\theta_A = \rho^\theta(y, A)/\rho^\theta(x, A)$. Thus $\succ^\theta$ and each δ^θ_A are revealed by ρ^θ.

It remains to show that the parameters $\Omega = \left\langle \pi(\theta_j), \rho^{\theta_j} \right\rangle_{j=1}^J$ continue to be generically identified when we restrict Equation 4 to be consistent with either of the two specialized models. To this end, we partition $\mathcal{A}$ as $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ such that $\mathcal{A}_3 = \{A^*\}$ with $|A^*| = 2$, $\mathcal{A}_1 = \{A \in \mathcal{A} : |A| = 3\}$, and $\mathcal{A}_2 = \mathcal{A} \setminus [\mathcal{A}_1 \cup \mathcal{A}_3]$. Enumerating $\mathbb{C}$ as $\langle c_i \rangle_{i=1}^I$, we write c_i^v for the restriction of c_i to partition element $\mathcal{A}_v$, and $\mathbb{C}^v$ for the set of all such restricted choice functions. For each v, define the matrix Φ^v with dimensions $I \times J$ and entry $\langle i, j \rangle$

equal to $\prod_{A\in\mathcal{A}_v} \rho^{\theta_j}(c_i^v(A), A)$. For the RST model this entry appears as

$$\Phi_{i,j}^v = \prod_{A\in\mathcal{A}_v} \left[\sum_{\tilde{x}\in A:\, c_i^v(A)\succeq^{\theta_j}\tilde{x}} \frac{\tau_A^{\theta_j}(\tilde{x})}{r_A^{\theta_j}(\tilde{x})} \right], \tag{34}$$

and for the QFC model as

$$\Phi_{i,j}^v = \prod_{A\in\mathcal{A}_v} [\delta_A^{\theta_j}]^{r_A^{\theta}(c_i^v(A))-1} \left[\frac{1-\delta_A^{\theta_j}}{1-[\delta_A^{\theta_j}]^{|A|}} \right]. \tag{35}$$

As in Allman et al. [4], we can establish identification of the RST model by proving that, for each $v = 1, 2$, the matrix defined in Equation 34 generically has full Kruskal rank. To show this, note first that the entries of this matrix are polynomial functions of the RST parameters $\tau_A^{\theta_j}(\tilde{x})$, and each of its minor determinants is thus also a polynomial in these parameters. Since a polynomial function defined on a Euclidean space is either identically zero or nonzero almost everywhere (see, e.g., Caron and Traynor [11]), it suffices to show that Φ^v has a minor determinant of order $n!$ which is nonzero for a single set of the RST parameters. This is demonstrated in Lemma 2 below.

The entries in Equation 35 are not polynomial functions of the QFC parameters $\delta_A^{\theta_j}$. However, we can multiply each column j of Φ^v by the scalar $\prod_{A\in\mathcal{A}_v}[1-[\delta_A^{\theta_j}]^{|A|}]$, without changing its rank, to obtain a modified matrix $\hat{\Phi}^v$ with polynomial entries

$$\hat{\Phi}_{i,j}^v = \prod_{A\in\mathcal{A}_v} [\delta_A^{\theta_j}]^{r_A^{\theta}(c_i^v(A))-1}[1-\delta_A^{\theta_j}]. \tag{36}$$

The argument in the previous paragraph then applies to this new matrix, with the required minor determinant and set of QFC parameters supplied by Lemma 2. □

Lemma 2. *For each $v = 1, 2$:* **(i)** *The matrix in Equation 34 has a minor determinant of order $n!$ that is nonzero for some RST parameters $\tau_A^{\theta_j}(\tilde{x})$.* **(ii)** *The matrix in Equation 36 has a minor determinant of order $n!$ that is nonzero for some QFC parameters $\delta_A^{\theta_j}$.*

Proof. There are four separate assertions in the lemma, corresponding to the two nontrivial partition elements $v = 1, 2$ times the two specialized models RST and QFC. All four cases proceed along the same lines, so for brevity we give a proof only for the (least notationally cumbersome) combination of $v = 1$ and the RST model.[27]

Let $v = 1$. Without loss of generality, we can renumber the $J = n!$ preference types so that, for each $j \leq J/2$, the order $\succ^{\theta_j}$ and its "partner" $\succ^{\theta_{j+J/2}}$ coincide except for over the two worst alternatives. Write $z^j(1) \succ^{\theta_j} z^j(2) \succ^{\theta_j} \cdots \succ^{\theta_j} z^j(n)$ for the ranking of type θ_j, and note that it agrees with its partner $\theta_{j+J/2}$ on the optimum from any menu of size at least three.

Again without loss of generality, we can now renumber the restricted choice functions in $\mathbb{C}^1$ such that the first J rows of the matrix Φ^1 correspond to the functions c_i^1 defined, for each $A \in \mathcal{A}_1$, by

$$c_i^1(A) = \begin{cases} \max_{\succ^{\theta_i}}(A) & \text{if either } i \leq J/2 \text{ or } A \neq z^i(1)z^i(2)z^i(3), \\ z^i(2) & \text{if both } J/2 < i \leq J \text{ and } A = z^i(1)z^i(2)z^i(3). \end{cases} \tag{37}$$

Observe that if $i \leq J/2 < i' \leq J$ then c_i^1 and $c_{i'}^1$ are distinct, since the first is rationalizable over $\mathcal{A}_1$ and the second is not. Moreover, if $i < i' \leq J/2$ then these functions are again distinct, since they are rationalized by orders that differ on rank positions above the two worst, and therefore choose differently from at least one menu in $\mathcal{A}_1$. Finally, if $J/2 < i < i' \leq J$ then c_i^1 and $c_{i'}^1$ are yet again distinct, since they must choose differently from at least one menu $A \in \mathcal{A}_1$ other than $z^i(1)z^i(2)z^i(3)$ or $z^{i'}(1)z^{i'}(2)z^{i'}(3)$. After the above renumberings, we denote by $\overline{\Phi}^1$ the square matrix consisting of the first J rows and columns of Φ^1.

Next we exhibit a choice of RST parameters such that the determinant of $\overline{\Phi}^1$ is nonzero, noting that this is an order J minor determinant of the original matrix Φ^1. For each $j \leq J/2$, we select a conditional SCF by requiring $\rho^{\theta_j}(\max_{\succ^{\theta_j}}(A), A) = 1$. Likewise, for

[27] Details for the other cases are available from the authors upon request.

$J/2 < j \leq J$ we require $\rho^{\theta_j}(\max_{\succ^{\theta_j}}(A), A) = 1$ when $A \neq z^j(1)z^j(2)z^j(3)$; as well as

$$\rho^{\theta_j}(z^j(1), z^j(1)z^j(2)z^j(3)) = 4/7, \tag{38}$$

$$\rho^{\theta_j}(z^j(2), z^j(1)z^j(2)z^j(3)) = 2/7. \tag{39}$$

These values can be obtained from the RST model by setting $\tau_A^{\theta_j}\left(\max_{\succ^{\theta_j}}(A)\right) = 1$ to yield $\rho^{\theta_j}(\max_{\succ^{\theta_j}}(A), A) = 1$ and $\tau^{\theta_j}_{z^j(1)z^j(2)z^j(3)}(z^j(1)) = \tau^{\theta_j}_{z^j(1)z^j(2)z^j(3)}(z^j(2)) = 2/7$ to yield Equations 38–39.

Evaluating $\overline{\Phi}^1$ for the conditional SCFs described in the previous paragraph, our goal is to show that the determinant of this $J \times J$ matrix is nonzero. To establish this, observe first that for $j \leq J/2$ the matrix entry $\prod_{A \in \mathcal{A}_1} \rho^{\theta_j}(c_i^1(A), A)$ is strictly positive if and only if $i = j$. Moreover, for $J/2 < j \leq J$ the same matrix entry is strictly positive if and only if either $i = j$ or $i = j - J/2$. These facts follow from the construction of $\overline{\Phi}^1$ (in particular Equation 37) and that of the conditional SCFs used to evaluate it.

Now let $\phi(j)$ denote the jth column of the matrix $\overline{\Phi}^1$, and suppose $\sum_{j=1}^J \xi_j \phi(j) = 0$ for scalars $\xi_1, \ldots, \xi_J$. The ith component of the latter equation states that

$$\sum_{j=1}^{J} \xi_j \prod_{A \in \mathcal{A}_1} \rho^{\theta_j}(c_i^1(A), A) = 0. \tag{40}$$

If $J/2 < i \leq J$ then we have that $\xi_i \prod_{A \in \mathcal{A}_1} \rho^{\theta_i}(c_i^1(A), A) = 0$, since the product vanishes in the other terms of the sum. But we also know that $\prod_{A \in \mathcal{A}_1} \rho^{\theta_i}(c_i^1(A), A) > 0$, and hence $\xi_i = 0$ in this case. If $i \leq J/2$, on the other hand, then by a similar logic we have

$$\xi_i \prod_{A \in \mathcal{A}_1} \rho^{\theta_i}(c_i^1(A), A) + \xi_{i+J/2} \prod_{A \in \mathcal{A}_1} \rho^{\theta_{i+J/2}}(c_i^1(A), A) = 0. \tag{41}$$

Once again $\prod_{A \in \mathcal{A}_1} \rho^{\theta_i}(c_i^1(A), A) > 0$, and here $\xi_{i+J/2} = 0$ since $i + J/2 > J/2$, so it follows that $\xi_i = 0$ in this case as well. We conclude that $\xi_1 = \xi_2 = \cdots = \xi_J = 0$, the columns of $\overline{\Phi}^1$ are linearly independent, and this matrix has nonzero determinant. $\square$

A.3 Practical formulation

In this section we sketch how our theoretical framework could be implemented in practice to estimate structural parameters from mixture choice data.

Enumerating the set $\mathbb{C}$ of choice functions as $\langle c_i \rangle_{i=1}^{I}$, we can write the mixture choice function μ in vector form as $\boldsymbol{\mu} = \langle \mu(c_i) \rangle_{i=1}^{I}$. Index a given collection of specialized models by $m = 1, \ldots, M$. Denote the SCF of preference/model type $\langle \succ^h, m \rangle$ by $\rho^{h,m}$, and write $\boldsymbol{\pi} = \langle \langle \pi(h,m) \rangle_{h=1}^{H} \rangle_{m=1}^{M}$ for the type distribution. Equation 4 can then be expressed as

$$\mu(c_i) = \sum_{m=1}^{M} \sum_{h=1}^{H} \pi(h,m) \prod_{k=1}^{K} \rho^{h,m}(c_i(A_k), A_k). \tag{42}$$

For each model m, the $I \times H$ matrix

$$\mathbf{R}(m) = \begin{bmatrix} \prod_{k=1}^{K} \rho^{1,m}(c_1(A_k), A_k) & \cdots & \prod_{k=1}^{K} \rho^{H,m}(c_1(A_k), A_k) \\ \vdots & & \vdots \\ \prod_{k=1}^{K} \rho^{1,m}(c_I(A_k), A_k) & \cdots & \prod_{k=1}^{K} \rho^{H,m}(c_I(A_k), A_k) \end{bmatrix} \tag{43}$$

describes the transition from the unobserved type distribution in $\langle \pi(h,m) \rangle_{h=1}^{H}$ to the observed mixture choice data. Lining up the M transition matrices to form the $I \times HM$ array $\mathbf{R}$=[$\mathbf{R}(1) \cdots \mathbf{R}(M)$], Equation 42 now becomes the vector equality $\boldsymbol{\mu} = \mathbf{R}\boldsymbol{\pi}$.

In the above scheme, the functions $\rho^{h,m}$ contain the deep parameters to be estimated. These are collected in a vector $\boldsymbol{\beta} = \langle \beta_\ell \rangle_{\ell=1}^{L}$, and we write $\rho_{\boldsymbol{\beta}}^{h,m}$ for a typical SCF to show dependence on the parameter assignments.

Next, consider an i.i.d. sample $Y = \langle Y_s \rangle_{s=1}^{S}$ of size S, where each $Y_s \in \mathbb{C}$ records the behavior of agent s over the menu collection $\langle A_k \rangle_{k=1}^{K}$. The probability of observing data point $Y_s = c_i$ is computed as $\mu(Y_s) = \mu(c_i) = \boldsymbol{\mu}_i = \mathbf{R}_i\boldsymbol{\pi}$, and the log-likelihood of the entire sample is then

$$\mathrm{LL}(Y|\boldsymbol{\pi}, \boldsymbol{\beta}) = \sum_{s=1}^{S} \log \left[\sum_{m=1}^{M} \sum_{h=1}^{H} \pi(h,m) \prod_{k=1}^{K} \rho_{\boldsymbol{\beta}}^{h,m}(Y_s(A_k), A_k) \right]. \tag{44}$$

Direct maximization of this function by choice of $\langle \boldsymbol{\pi}, \boldsymbol{\beta} \rangle$ is possible in theory, but may be computationally challenging. An alternative is to employ the Expectation Maximization (EM) algorithm, at each step imposing constraints appropriate to the specialized model at hand (e.g., each $\tau^h_A(\tilde{x}) \in [0,1]$ for the RST model).[28]

References

[1] Jason Abaluck and Abi Adams. What do consumers consider before they choose? Identification from asymmetric demand responses. *Quarterly Journal of Economics*, forthcoming.

[2] Victor H. Aguiar, Maria Jose Boccardi, and Mark Dean. Satisficing and stochastic choice. *Journal of Economic Theory*, 166:445–482, November 2016.

[3] Victor H. Aguiar, Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim. Does random consideration explain behavior when choice is hard? Evidence from a large-scale experiment. Working paper (arXiv:1812.09619), June 2019.

[4] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics*, 37(6A):3099–3132, December 2009.

[5] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. The dynamics of repeat consumption. *Proceedings of the 23rd International Conference on the World Wide Web*, pages 419–429. April 2014.

[28] Our framework amounts to a latent class model, where the response variables (i.e., the choices from each A_k) are independent conditional on the class (i.e., the preference/model type $\langle \succ^h, m \rangle$). Maximum likelihood estimation of such models can be carried out using the EM algorithm, a computational approach with a long tradition in statistics (see, e.g., McLachlan and Peel [32]). Here the log-likelihood expression in Equation 44 is maximized by alternating the following two steps until convergence is achieved: [1.] Given the observed choices Y and the current parameter estimates $\hat{\boldsymbol{\beta}}$, compute the expectation of log-likelihood with respect to the current conditional type distribution $\hat{\boldsymbol{\pi}}$. [2.] Update the parameter estimates in $\hat{\boldsymbol{\beta}}$ by maximizing the expected log-likelihood found in the first step.

[6] Jose Apesteguia, Miguel A. Ballester, and Angelo Gutierrez-Daza. Random models for the joint treatment of risk and time preferences. Working paper, March 2020.

[7] Jose Apesteguia, Miguel A. Ballester, and Jay Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, March 2017.

[8] Kenneth J. Arrow. Rational choice functions and orderings. *Economica*, New Series, 26(102):121–127, May 1959.

[9] Levon Barseghyan, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, forthcoming.

[10] H. D. Block and Jacob Marschak. Random orderings and stochastic theories of response. In Ingram Olkin, Sudhist G. Ghurye, Wassily Hoeffding, William G. Madow, and Henry B. Mann, editors. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 97–132. Stanford University Press, Stanford CA, 1960.

[11] Richard Caron and Tim Traynor. The zero set of a polynomial. Working paper, Department of Mathematics and Statistics, University of Windsor, 2005.

[12] Matias D. Cattaneo, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. A random attention model. *Journal of Political Economy*, 128(7):2796–2836, July 2020.

[13] Giovanni Compiani and Yuichi Kitamura. Using mixtures in econometric models: A brief review and some new results. *Econometrics Journal*, 19(3):C95–C127, October 2016.

[14] Valentino Dardanoni, Paola Manzini, Marco Mariotti, and Christopher J. Tyson. Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, 88(3):1269–1296, May 2020.

[15] Rohan Dutta and Sean Horan. Inferring rationales from choice: Identification for rational shortlist methods. *American Economic Journal: Microeconomics*, 7(4):179–201, November 2015.

[16] Gustav T. Fechner. *Elements of Psychophysics*. Helmut E. Adler, translator; Davis H. Howes and Edwin G. Boring, editors. New York: Holt, Rinehart and Winston, 1966.

[17] Emel Filiz-Ozbay and Yusufcan Masatlioglu. Progressive random choice. Working paper, April 2020.

[18] Peter C. Fishburn. Stochastic utility. In Salvador Barbera, Peter J. Hammond, and Christian Seidl, editors, *Handbook of Utility Theory*, pages 273–318. Boston: Kluwer Academic, 1998.

[19] Glenn W. Harrison and E. Elisabet Rutström. Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, 12(2):133–158, June 2009.

[20] Sheena S. Iyengar and Mark R. Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6):995–1006, December 2000.

[21] Nail Kashaev and Victor H. Aguiar. A random attention and utility model. Working paper (arXiv:2105.11268), May 2021.

[22] Matthew Kovach and Gerelt Tserenjigmid. Behavioral foundations of nested stochastic choice and nested logit. Working paper, November 2019.

[23] Joseph B. Kruskal (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications* 18:95–138.

[24] Arthur Lewbel. The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903, December 2019.

[25] Yi-Hsuan Lin. Random non-expected utility: Non-uniqueness. Working paper (arXiv:2009.04173), September 2020.

[26] Juan Sebastián Lleras, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. When more is less: Limited consideration. *Journal of Economic Theory*, 170:70–85, July 2017.

[27] R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley, 1959.

[28] Paola Manzini and Marco Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176, May 2014.

[29] Paola Manzini and Marco Mariotti. State dependent choice. *Social Choice and Welfare*, 45(2):239–268, September 2015.

[30] Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. Revealed attention. *American Economic Review*, 102(5):2183–2205, August 2012.

[31] Daniel L. McFadden. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press: New York, 1974.

[32] Geoffrey McLachlan and David Peel. *Finite Mixture Models*. New York: Wiley, 2000.

[33] Douglas A. McManus. How common is identification in parametric models? *Journal of Econometrics*, 53(1–3):5–23, July–September 1992.

[34] Yuval Salant and Ariel Rubinstein. (A, f): Choice with frames. *Review of Economic Studies*, 75(4):1287–1296, October 2008.

[35] Paul A. Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, New Series, 5(17):61–71, February 1938.

[36] Michelle Sovinsky Goeree. Limited information and advertising in the U.S. personal computer industry. *Econometrica*, 76(5):1017–1074, September 2008.

[37] Tomasz Strzalecki. Stochastic choice. Hotelling Lectures in Economic Theory, August 2017.

[38] Louis L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, July 1927.

[39] Christopher Turansick. Identification in the random utility model. Working paper (arXiv:2102.05570), April 2021.

[40] Christopher J. Tyson. Exponential satisficing. *American Economic Journal: Microeconomics*, 13(2):439–467, May 2021.

[41] Andrea Wilson. Bounded memory and biases in information processing. *Econometrica*, 82(6):2257–2294, November 2014.