Mixture Choice Data: Revealing Preferences and Cognition*

Valentino Dardanoni  Paola Manzini
Università degli Studi di Palermo  University of Sussex and IZA

Marco Mariotti  Henrik Petri
Queen Mary University of London  University of Bath

Christopher J. Tyson
Queen Mary University of London

June 1, 2020

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 cognition, we show that these two components of behavior can be revealed simultaneously by appropriate mixture choice data. We extend this finding to several models of cognition, including stochastic consideration sets, random satisficing thresholds, and multinomial logit. Finally, we demonstrate how the mixture choice framework can be used by applying it to an experimental dataset.

J.E.L. classification codes: D01, D91.

Keywords: attention, heterogeneity, identification, satisficing, stochastic choice.

*We would like to thank Emel Filiz-Ozbay, Guillaume Fréchette, Yusufcan Masatlioglu, and a seminar audience at ULB (ECARES) for valuable comments and suggestions.
1 Introduction

Although economic orthodoxy has always viewed preferences as the primary determinant of individual choice behavior, it has long been acknowledged that other cognitive factors can play a significant role. The decision maker’s working memory, cognitive load, mood or affect, framing of the situation, ingrained biases and heuristics, degree of attention to the environment, and general intelligence can all be shown to influence choice in ways that are difficult to explain under the preference maximization paradigm. Moreover, just as Samuelson [27], Arrow [7], and their successors elucidated the behavioral consequences of preference-based choice, the past two decades have seen a profusion of theoretical work on the observable implications of various types of cognitive constraints for decision making at the individual level.\footnote{This literature is too large to summarize here, but some early contributions include Kalai et al. [17], Manzini and Mariotti [22], Tyson [31], Salant and Rubinstein [26], and Masatlioglu et al. [24].}

A separate strand of research in decision theory, originating with Luce [20] and Block and Marschak [9], focuses on the problem of relating heterogeneous individual preferences to aggregate choice data. This has led to a highly productive interplay with econometrics, and to the development of discrete choice models that have become essential tools in labor economics, industrial organization, and other fields. However, cognitive factors were until recently for the most part ignored by this research program, leaving open the challenge of understanding the aggregate behavior of agents who are not fully reliable preference maximizers.\footnote{Apesteguia et al. [5, 6] and Kovach and Tserenjigmid [18] exemplify recent theoretical work advancing this agenda.}

Following this train of inquiry, there is no compelling reason to expect a population of decision makers to be more homogeneous in their cognitive traits and abilities than we would expect them to be in their preferences. Ideally, we would wish to allow heterogeneity in both dimensions—and even correlation between them—while continuing to work with aggregate data. But many existing models with variable cognition either are formulated in a single-agent setting (e.g., Dutta and Horan [13] and Lleras et al. [19]) or assume that preferences are uniform (e.g., Manzini and Mariotti [23] and Aguiar et al. [2]). Of course, it is straightforward enough to write down a model accommodating “double heterogeneity”
of the desired sort. The difficulty here is constructing a framework that can combine this degree of flexibility with attractive identification properties, enabling us to separate out preferences and cognition and to pin down the parameters of each component.

A few recent papers have tackled this problem directly, including Sovinsky Goeree [28], Abaluck and Adams [1], Aguiar et al. [3], and Barseghyan et al. [8]. These contributions either pursue partial identification or gain the leverage needed for exact identification via individual covariates (e.g., demographics), alternative covariates (e.g., advertising levels), or experimental control of cognition. In contrast to this work, we aim to develop a general methodology that remains entirely choice-theoretic (avoiding the use of covariates) and that can identify doubly heterogeneous models capturing various aspects of cognition.

The proposed methodology takes as its starting point “mixture choice data” consisting of the joint distribution of choices of a population of agents from a collection of menus. For example, in an experiment where subjects are asked to choose from several different menus of options, the researcher will know not only the distribution of choices from each menu separately—the empirical stochastic choice function—but also the proportion of subjects who make a given combination of choices across the collection of problems. On a much larger scale, 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 will be a probability distribution over deterministic choice functions (i.e., maps from menus to available choices), interpreted as returning the share of the population observed to choose in a particular way across the relevant domain of problems. As noted, a mixture choice function contains richer information than the corresponding stochastic choice function, since it records correlations among the choices from different menus. We will find that knowing these cross-menu correlations is essential for our goal of disentangling the preferential and cognitive components of behavior.\(^3\)

\(^3\)Filiz-Ozbay and Masatlioglu [15] simultaneously and independently propose the concept of a mixture
As in a random utility model, we posit a heterogeneous population of decision makers with different “preference types”; that is, preference orders over the alternatives. However, the individuals of a given type do not consistently maximize these preferences, since their choices can also be affected by cognitive factors such as mood, framing, and attention. These factors are realized stochastically and independently within each choice problem in the dataset, while the agent’s preferences remain stable across all problems. The observed shares of choice patterns are therefore generated by an unobserved mixture (over types) of unobserved type-conditional stochastic choices from each menu.4

For the doubly heterogeneous model described above, exact identification requires that the decomposition of our mixture choice dataset into type-conditional stochastic choice functions be unique. When this holds, the observed data will tell us each preference type’s share of the population as well as its distribution of choices from each menu after cognitive factors wield their influence. Of course, identification does not come for free: Instead, we offer a variety of results (namely, Propositions 1–4 and Corollaries 1–3) which show that the desired uniqueness can be guaranteed generically, but only if the menus included in our dataset are large and numerous enough relative to the number of preference types.5

Identification results of this sort possess inherent interest, but may also be useful for experimental design. Suppose, for instance, that one wishes to determine the distribution of preferences in a population of boundedly rational agents via a choice experiment. 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 collection of menus presented? Our findings are well suited to address questions such as these.

choice function, focusing on its relationship to the corresponding stochastic choice function (see Section 2.1 below). These authors prove a number of results concerning the representation of stochastic choice data by mixture choice functions with structured support. Our contribution, on the other hand, takes mixture choice data to be observable and focuses on the identification of preference and cognitive parameters. In this sense, the two papers can be seen as complementary.

4A more formal expression of the structure we impose on mixture choice data appears in Equation 3. 5These identification results rely on a powerful theorem from Allman et al. [4], adapted as Lemma 1. Related techniques from the statistical literature were used in Dardanoni et al. [12] to study the converse situation, where cognitive characteristics are assumed to be stable across choice problems but preferences are realized stochastically. Among other differences with the present paper, the analysis in [12] centered on a cognitive model (of agents with limited “consideration capacity”) that we do not examine here.
Having investigated sufficient conditions for identification in the general mixture choice framework, we then extend the analysis to a number of specialized models of cognition. More precisely, we study stochastic consideration sets, random satisficing thresholds, the multinomial logit model, and a broad class of Fechnerian models. The task here is to use our knowledge of each preference type’s stochastic choice function (supplied by an earlier identification result) to pin down the “deep” cognitive parameters of the specialized model at hand. We give conditions for identification under the stochastic consideration, random satisficing, and multinomial logit models (Propositions 5–7); thus showing that mixture choice data allow us to reveal any of these forms of bounded rationality in the presence of preference heterogeneity. Notably, these results permit substantial correlation between preferences and cognition, as well as menu-dependence of many cognitive parameters.

Finally, we sketch an indicative, “proof-of-concept” application of our methodology to a preexisting mixture choice dataset. Using the behavior of subjects in a time preference experiment, we estimate (by maximum likelihood) the stochastic consideration sets model, the random satisficing thresholds model, and a “quantal” model from the Fechnerian class. We compare the estimated distributions of preference types yielded by these three models, and examine their relative performance according to familiar model selection criteria.

2 The general mixture choice model

2.1 Stochastic and mixture choice functions

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 such menus. We refer to the set $\{A \subseteq X : |A| \geq 2\}$ as the full collection of menus.

A choice function is any rule $c : \mathcal{A} \rightarrow X$ 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 $C$; i.e., a $\mu : C \to [0, 1]$ such that $\sum_{c \in C} \mu(c) = 1$. The MCF $\mu$ is said to be menu independent if $\forall c \in C$ we have $\mu(c) = \prod_{A \in \mathcal{A}} \sum_{c' \in C : c'(A) = c(A)} \mu(c')$.

Menu independence eliminates interaction across choice problems, returning us to the SCF framework. To see this, note that any SCF $\rho$ induces the (“product”) MCF defined by $\mu_{\rho}(c) = \prod_{A \in \mathcal{A}} \rho(c(A), A)$, and conversely any MCF $\mu$ can be used to construct the (“marginal”) SCF $\rho_{\mu}(x, A) = \sum_{c \in C : c(A) = x} \mu(c)$. The latter procedure will preserve all of the information in the MCF if and only if the probability of each choice function is the product across menus of the probabilities of the chosen alternatives—which is precisely the menu-independence property.

Observation 1. The mapping $\rho \mapsto \mu_{\rho}$ is a bijection from the set of all stochastic choice functions onto the set of menu-independent mixture choice functions, with inverse $\mu \mapsto \rho_{\mu}$.

To confirm this equivalence, note that for any SCF $\rho$ and $x \in A \in \mathcal{A}$ we have

$$\rho_{\mu_{\rho}}(x, A) = \sum_{c \in C : c(A) = x} \left[ \prod_{B \in \mathcal{A}} \rho(c(B), B) \right] = \rho(x, A) \sum_{c \in C : c(A) = x} \left[ \prod_{A \neq B \in \mathcal{A}} \rho(c(B), B) \right].$$ (1)

On the other hand, given a menu-independent MCF $\mu$ and a $c \in C$, we have that $\mu_{\rho_{\mu}}(c) = \prod_{A \in \mathcal{A}} \sum_{c' \in C : c'(A) = c(A)} \mu(c') = \mu(c)$. For instance, if $\rho(x, xy) = r_1$ and $\rho(x, xz) = r_2$, then $\mu_{\rho}(c(xy) = c(xz) = x) = r_1 r_2$, $\mu_{\rho}(c(xy) = x \land c(xz) = z) = r_1 [1 - r_2]$, and so on.

Note that the above equivalence relies on mixture choice functions being defined over the entire space $C$, and may not hold on smaller spaces. For instance, consider the class $C^r \subset C$ of choice functions $c$ that are rational in the sense that there exists a linear order (i.e., a complete, transitive, and antisymmetric relation) $\succeq$ on $X$ such that $\forall x \in A \in \mathcal{A}$ we have $c(A) \succeq x$. If $\mu$ is an MCF whose support is a subset of $C^r$, then

$$\rho_{\mu}(x, xyz) = \sum_{c \in C^r : c(xyz) = x} \mu(c) \leq \sum_{c \in C : c(xy) = x} \mu(c) = \rho_{\mu}(x, xy),$$ (2)

which rules out the construction of any SCF $\rho$ with $\rho(x, xyz) > \rho(x, xy)$. In view of this
example, it is clear that MCFs restricted to $\mathbb{C}^r$ can only generate SCFs that are consistent with random utility maximization (as defined by Block and Marschak [9]), and not the entire space of possible SCFs.\footnote{For an analysis of structured MCFs, see Filiz-Ozbay and Masatlioglu [15].}

## 2.2 Preference types and aggregate choices

In this paper, our primary objective is to use mixture choice functions as tools to achieve simultaneous identification of preferential and cognitive factors affecting choice behavior, when agents are heterogeneous on both dimensions. Specifically, we imagine a population composed of individuals with one of a finite number of preference types $\theta \in \Theta$. These types could represent either full preference orders over $X$ or values of a (discretized) numerical preference parameter such as the agent’s discount factor or coefficient of risk aversion. The type distribution is denoted by $\langle \pi(\theta) \rangle_{\theta \in \Theta}$, and each type $\theta$ has a stochastic choice function $\rho^\theta$ realized independently across the menus in $\mathcal{A}$. We assume that the researcher observes the mixture choice function

$$
\mu(c) = \sum_{\theta \in \Theta} \pi(\theta) \mu_{\rho^\theta}(c) = \sum_{\theta \in \Theta} \pi(\theta) \prod_{A \in \mathcal{A}} \rho^\theta(c(A), A), \tag{3}
$$

formed as the expectation over $\theta$ of the product MCF induced by $\rho^\theta$. More explicitly, the likelihood of choice function $c \in \mathbb{C}$ is the probability-weighted sum of the likelihood that each type $\theta$ independently chooses alternative $c(A)$ from each menu $A$.

Several points about Equation 3 are worth noting. Firstly, the determination of $\mu(c)$ involves two nested levels of stochasticity: Conditional on $\theta$, the agent’s SCF $\rho^\theta$ is random due to unobserved cognitive variables (such as mood, framing, or attention), which we take to be statistically independent across menus. Moreover, from the researcher’s perspective the type $\theta$ is also random, since the agent’s preferences are unobserved. In summary, we are uncertain about the pattern of choices that will arise both because we have imperfect knowledge of tastes and because these tastes do not mechanically predict behavior.

Secondly, the assumption of independence across menus conditional on type is not in general inherited by the observed aggregate (i.e., unconditional) choices. While the MCF
µ in Equation 3 leads to the SCF \( \rho_\mu(x, A) = \sum_{\theta \in \Theta} \pi(\theta) \rho^\theta(x, A) \), this marginal information cannot typically be used to reconstruct the aggregate behavior. Indeed, if we in turn use \( \rho_\mu \) to induce the MCF \( \mu_\rho(c) = \prod_{A \in \mathcal{A}} \sum_{\theta \in \Theta} \pi(\theta) \rho^\theta(c(A), A) \), which forces independence, then the result will differ from the original \( \mu \) (unless, by coincidence, the latter MCF was menu-independent in the first place).

Thirdly, no feature of Equation 3 requires \( \theta \) to represent a preference type and \( \rho^\theta \) to represent cognitive factors. We adopt this interpretation for concreteness and because we view the present framework as well suited to applications along these lines. But the roles of preferences and cognition could be reversed (see Footnote 5), or either component could correspond to some third category of influences on decision making. Note that, under any interpretation, our central assumption remains statistical independence of choice behavior across menus conditional on the agent’s type.

For notational convenience, we enumerate the collection \( \mathcal{A} \) as \( \langle A_1, \ldots, A_K \rangle \) and denote by \( \langle x_1, \ldots, x_K \rangle \) the choice function \( c \) with each \( c(A_k) = x_k \).

**Example 1.** Let the collection \( \mathcal{A} \) consist of the binary menus drawn from \( X = xyz \), enumerated as \( \langle xy, xz, yz \rangle \). Suppose further that \( \Theta = \{ \theta_1, \theta_2 \} \), with \( \pi(\theta_1) = \pi(\theta_2) = 1/2 \), and write \( u^\theta : X \to \mathbb{R} \) for the utility function of type \( \theta \). Assume that this type chooses \( w \) from \( A \) with the multinomial logit probability

\[
\rho^\theta(w, A) = \frac{\exp u^\theta(w)}{\sum_{w' \in A} \exp u^\theta(w')}; \quad (4)
\]

and set \( u^{\theta_1}(x) = u^{\theta_2}(z) = \log 3 \), \( u^{\theta_1}(y) = u^{\theta_2}(x) = \log 2 \), and \( u^{\theta_1}(z) = u^{\theta_2}(y) = \log 1 = 0 \).
Equation 3 then yields the MCF $\mu$ defined by

$$\mu(x, x, y) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][2/5][1/4] = 22/120 \approx 0.183,$$

(5)

$$\mu(x, x, z) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][2/5][3/4] = 21/120 = 0.175,$$

(6)

$$\mu(x, z, y) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][3/5][1/4] = 12/120 = 0.100,$$

(7)

$$\mu(x, z, z) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][3/5][3/4] = 21/120 = 0.175,$$

(8)

$$\mu(y, x, y) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][2/5][1/4] = 14/120 \approx 0.117,$$

(9)

$$\mu(y, x, z) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][2/5][3/4] = 12/120 = 0.100,$$

(10)

$$\mu(y, z, y) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][3/5][1/4] = 7/120 \approx 0.058,$$

(11)

$$\mu(y, z, z) = \frac{1}{2}[2/3][3/5][1/3] + [1/2][2/3][3/5][3/4] = 11/120 \approx 0.092.$$

(12)

The associated marginal SCF $\rho_{\mu}$ is given by $\rho_{\mu}(x, xy) = 19/30$, $\rho_{\mu}(x, xz) = 23/40$, and $\rho_{\mu}(y, yz) = 11/24$. These choice probabilities in turn induce the MCF $\mu_{\rho_{\mu}}$ given by

$$\mu_{\rho_{\mu}}(x, x, y) = \frac{19}{30}[23/40][11/24] = 4807/28800 \approx 0.167,$$

(13)

$$\mu_{\rho_{\mu}}(x, x, z) = \frac{19}{30}[23/40][13/24] = 5681/28800 \approx 0.197,$$

(14)

and so on; which can be interpreted as the projection of the function $\mu$ in Equations 5–12 onto the subspace of menu-independent MCFs. ||

3 Generic identification results: A tool kit

3.1 Sufficient conditions for identification

We are interested in the identification properties of the general mixture choice model in Equation 3. In particular, our focus is on the number $J = |\Theta|$ of preference types that the model allows subject to the condition that its parameters are uniquely determined by the observable data. In Section 3 we address this issue without imposing any further structure on the model; that is, without making any assumptions about the cognitive factors that produce deviations from preference maximization. Here we are content to use the mixture
choice data to identify the distributions of types and of type-conditional choices, while in Section 4 we will proceed to examine specialized models of bounded rationality, for which the identification of “deep” cognitive parameters becomes possible.

Enumerating the types in $\Theta$ as $\langle \theta_1, \ldots, \theta_J \rangle$, the parameters of the model in Equation 3 appear as $\Omega = \langle \pi(\theta_j), \rho^{\theta_j} \rangle_{j=1}^J$, where each $\pi(\theta_j)$ is the probability of type $\theta_j$ and $\rho^{\theta_j}$ is the corresponding SCF. This model is (strictly) identified if the mapping $\Omega \mapsto \mu$ is one-to-one, and generically identified if the same property holds except on a set of parameter values of (Lebesgue) measure zero. Note that within Section 3 we ignore “label swapping”; namely, modifying $\Omega$ only by changing the arbitrary enumeration of $\Theta = \{\theta_1, \ldots, \theta_J\}$, which leaves the content of the model entirely unaffected. In more specific models the types will carry a concrete interpretation—such as higher or lower rationality in some sense—and we will wish to know which label $\theta_j$ corresponds to which substantive variety of agent. We will see in Section 4 how type labels are thus “pinned down.”

To address the question of (generic) identification we shall use sufficient conditions from Allman et al. [4, Theorem 4], adapted for our purposes as follows.

**Lemma 1** (Allman et al. [4]). Let $K \geq 3$ and suppose that $A_1, A_2, A_3 \subset A$ are disjoint and nonempty, with $A_1 \cup A_2 \cup A_3 = A$. Writing $\kappa_i = \prod_{A \in A_i} |A|$, we then have

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

(15)

only if the model is generically identified.

This feature of the model is illustrated in our next example.

**Example 2.** Let the collection $A$ consist of the binary menus drawn from $X = xyz$, enumerated as $\langle xy, xz, yz \rangle$. Temporarily adopting the notation $p = \pi(\theta_1), a_j = \rho^{\theta_j}(x, xy), \ldots$
\[ b_j = \rho^j(x, xz), \text{ and } d_j = \rho^j(y, yz); \text{ for } J = 2 \text{ we have } \]

\[
\mu(x, x, y) = p a_1 b_1 d_1 + [1 - p] a_2 b_2 d_2, \quad (16)
\]

\[
\mu(x, x, z) = p a_1 b_1 [1 - d_1] + [1 - p] a_2 b_2 [1 - d_2], \quad (17)
\]

\[
\mu(x, z, y) = p a_1 [1 - b_1] d_1 + [1 - p] a_2 [1 - b_2] d_2, \quad (18)
\]

\[
\mu(x, z, z) = p a_1 [1 - b_1] [1 - d_1] + [1 - p] a_2 [1 - b_2] [1 - d_2], \quad (19)
\]

\[
\mu(y, x, y) = p [1 - a_1] b_1 d_1 + [1 - p] [1 - a_2] b_2 d_2, \quad (20)
\]

\[
\mu(y, x, z) = p [1 - a_1] b_1 [1 - d_1] + [1 - p] [1 - a_2] b_2 [1 - d_2], \quad (21)
\]

\[
\mu(y, z, y) = p [1 - a_1] [1 - b_1] d_1 + [1 - p] [1 - a_2] [1 - b_2] d_2, \quad (22)
\]

\[
\mu(y, z, z) = p [1 - a_1] [1 - b_1] [1 - d_1] + [1 - p] [1 - a_2] [1 - b_2] [1 - d_2]. \quad (23)
\]

Letting each \( \mathcal{A}_i = \{A_i\} \) yields \( \kappa_1 = \kappa_2 = \kappa_3 = 2 \), so that the condition in Equation 15 holds with equality; viz. \( 2 + 2 + 2 \geq 2 \times 2 + 2 \). Accordingly, the multi-linear system in Equations 16–23 (any seven of which are mutually independent) is generically solvable for the parameters \( \langle p, a_1, b_1, d_1, a_2, b_2, d_2 \rangle \).

For instance, substituting the mixture choice probabilities in Equations 5–12 into the left-hand-sides of (any seven of) Equations 16–23, we obtain a multi-linear system with unique solution \( \langle p, a_1, b_1, d_1, a_2, b_2, d_2 \rangle = (1/2, 3/5, 3/4, 2/3, 2/3, 2/5, 1/4) \). These are, of course, the same parameters used in Example 1 to construct the original MCF \( \mu \).

### 3.2 How many preference 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 \( \mathcal{J}(\mathcal{A}) \) 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 value of \( J \) that satisfies the inequality in Equation 15.

Throughout Section 3, we assume without loss of generality that \(|A_1| \geq \cdots \geq |A_K|\). Clearly both \(|A_1| \leq n \) and \(|A_K| \geq 2\), but neither of these constraints necessarily holds

---

\( ^7 \)As noted, for the moment we ignore label-swapping; which here would mean writing the parameters (equivalently) as \( (p', a'_1, b'_1, d'_1, a'_2, b'_2, d'_2) = (1/2, 2/3, 2/5, 1/4, 3/5, 3/4, 2/3) = (1 - p, a_2, b_2, d_2, a_1, b_1, d_1) \).
with equality in the absence of hypotheses about the content of \( \mathcal{A} \). In Equation 15, any partition that achieves \( \mathcal{J}(\mathcal{A}) \) will assign a menu with the smallest available cardinality to its own element, which we express as \( \mathcal{A}_3 = \{ A_K \} \). Since \( \kappa_3 = |A_K| \geq 2 \), the inequality in question is then implied by \( \min\{\kappa_1, \kappa_2\} \geq \mathcal{J}(\mathcal{A}) \), and we can write

\[
\mathcal{J}(\mathcal{A}) = \max_{\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{A} \setminus \mathcal{A}_3} \min\{\kappa_1, \kappa_2\}
\]

for the resulting upper bound on the number of types.

Ideally we would partition \( \mathcal{A} \setminus \mathcal{A}_3 \) so as to set \( \kappa_1 = \kappa_2 \), enabling the upper bound \( \mathcal{J}(\mathcal{A}) \) to attain its “theoretical optimum” value

\[
\mathcal{J}^*(\mathcal{A}) = \left[ \prod_{k=1}^{\left\lfloor \frac{K-1}{2} \right\rfloor} |A_k| \right]^{1/2}
\]

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

We adopt a constructive approach to this problem, making use of specific partitioning algorithms to approximate \( \mathcal{J}(\mathcal{A}) \) under various assumptions about the menus in \( \mathcal{A} \). With no such assumptions at all, we can bound \( \mathcal{J}(\mathcal{A}) \) from below by means of a simple partition that places the even-indexed menus (\( A_2, A_4, \) etc.) in \( \mathcal{A}_2 \) and the odd-indexed menus (\( A_1, A_3, \) etc.) in \( \mathcal{A}_1 \). Thus we obtain a baseline result that gauges the “identification power” of an arbitrary collection \( \mathcal{A} \) by calculating the product of the cardinalities of the menus with even indices (excluding \( A_K \)).

**Proposition 1.** If \( K \geq 3 \) and \( J \leq \prod_{k=1}^{\left\lfloor \frac{K-1}{2} \right\rfloor} |A_{2k}| \), then the model is generically identified.

**Proof.** Let \( \mathcal{A}_3 = \{ A_K \} \), \( \mathcal{A}_2 = \{ A_{2k} : 1 \leq k \leq \left\lfloor K - 1 \right\rfloor/2 \} \) (with \( k \) constrained to assume integer values), and \( \mathcal{A}_1 = \mathcal{A} \setminus \{ A_2 \cup A_3 \} \). Since \( |A_{2k-1}| \geq |A_{2k}| \) for \( k \leq \left\lfloor K - 1 \right\rfloor/2 \), with each menu in \( \mathcal{A}_2 \) is associated a menu in \( \mathcal{A}_1 \) with weakly higher cardinality. This ensures that \( \kappa_1 \geq \kappa_2 \), and it follows that \( \mathcal{J}(\mathcal{A}) \geq \min\{\kappa_1, \kappa_2\} = \kappa_2 = \prod_{k=1}^{\left\lfloor \frac{K-1}{2} \right\rfloor} |A_{2k}| \), as desired.\[\Box\]

---

8 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.

9 Note that here the smallest unassigned menu \( A_{K-1} \) is in \( \mathcal{A}_1 \) for \( K \) even and in \( \mathcal{A}_2 \) for \( K \) odd.
Despite the crudeness of the menu partition leading to Proposition 1, this result may be useful to establish identification in settings where the collection $\mathcal{A}$ must be taken as given and there is not too much variation in menu cardinalities. The following example illustrates this point in an environment where a specific alternative is on each menu.

**Example 3.** Let $X = wxyz$ and $\mathcal{A} = \{A \subseteq X : w \in A \land |A| \geq 2\}$; i.e. all menus contain alternative $w$. Here the list of cardinalities appears as $\langle |A_1|, \ldots, |A_7| \rangle = \langle 4, 3, 3, 3, 2, 2, 2 \rangle$, and the partition used in Proposition 1 leads to $\kappa_3 = |A_7| = 2$, $\kappa_2 = |A_2| \cdot |A_4| \cdot |A_6| = 18$, and $\kappa_1 = |A_1| \cdot |A_3| \cdot |A_5| = 24$. Thus generic identification is ensured with up to $\kappa_2 = 18$ types included in the model. Moreover, observe that in this case the specified partition solves the maximization problem in Equation 24, since no value closer to the theoretical optimum $J^*(\mathcal{A}) = \sqrt[4]{432} \approx 20.8$ is achievable. Hence we conclude that $J(\mathcal{A}) = 18$. ||

While Example 3 shows that the even/odd partition used in Proposition 1 can attain the optimum $J^*(\mathcal{A})$, it does not do so in general and hence our identification exercise can benefit from more sophisticated (albeit more laborious) partitioning methods. To produce tighter bounds, we will employ a “greedy partition” that assigns the menus in the sequence $\langle A_1, \ldots, A_{K-1} \rangle$ 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. This partition can be defined more formally as follows.

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

\begin{align*}
\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} \\
\mathcal{B}_2(k+1) = \{A_1, \ldots, A_{k+1}\} \setminus \mathcal{B}_1(k+1).
\end{align*}

The greedy partition of $\mathcal{A}$ then has $\mathcal{A}_1 = B_1(K-1)$, $\mathcal{A}_2 = B_2(K-1)$, and $\mathcal{A}_3 = \{A_K\}$.

The greedy partition is most effective when there are no large downward jumps in the sequence $|A_1| \geq \cdots \geq |A_K|$ of menu cardinalities. Our next example illustrates how this feature is present in one empirically relevant scenario.
Example 4. Consider “leave-one-out” menu variation, as used by Abaluck and Adams [1] to identify parameters in an econometric specification of limited attention models. More precisely, for \( n = |X| \geq 5 \) let \( \mathcal{A} \) contain both \( X \) itself and each menu obtained from \( X \) by removing one alternative. Here \( K = n + 1, |A_1| = n, \) and \( |A_2| = \cdots = |A_{n+1}| = n - 1. \)

When \( n = 5 \) and hence \( K = 6 \), the greedy algorithm sets \( \mathcal{A}_3 = \{A_6\} \) and constructs

\[
\begin{align*}
\mathcal{B}_1(1) &= \mathcal{B}_1(2) = \mathcal{B}_1(3) = \{A_1\} \subset \mathcal{B}_1(4) = \mathcal{B}_1(5) = \{A_1, A_4\}, \\
\mathcal{B}_2(1) &= \emptyset \subset \mathcal{B}_2(2) = \{A_2\} \subset \mathcal{B}_2(3) = \mathcal{B}_2(4) = \{A_2, A_3\} \subset \mathcal{B}_2(5) = \{A_2, A_3, A_5\};
\end{align*}
\]

(28)

yielding \( \mathcal{A}_1 = \{A_1, A_4\} \) and \( \mathcal{A}_2 = \{A_2, A_3, A_5\} \). Accordingly, \( \kappa_1 = |A_1| \cdot |A_4| = 5 \cdot 4 = 20 \), \( \kappa_2 = |A_2| \cdot |A_3| \cdot |A_5| = 4^3 = 64 \), and we find that up to \( \min\{\kappa_1, \kappa_2\} = \kappa_1 = 20 \) types can be included in the model while ensuring generic identification. In fact, no better partition is available in this case, and so we have \( J(\mathcal{A}) = 20. \)^{10}

Moving to the general case—and anticipating the manner of reasoning used in later results—we can confirm that after assigning \( A_1 = X \) to \( \mathcal{A}_1 \), the greedy partition assigns two menus of size \( n - 1 \) to \( \mathcal{A}_2 \) and then the remaining \( K - 4 = n - 3 \) menus to \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) in an alternating fashion. If \( n \) is odd (and hence \( K \) is even), then \( \mathcal{A}_1 \) ends up containing \( \lceil K - 4 \rceil / 2 \) menus of size \( n - 1 \), with \( \mathcal{A}_2 \) containing \( 2 + \lceil K - 4 \rceil / 2 = K / 2 \) such menus. On the other hand, if \( n \) is even (and hence \( K \) is odd) then the menus of size \( n - 1 \) are assigned \( 1 + \lceil K - 5 \rceil / 2 = \lceil K - 3 \rceil / 2 \) to \( \mathcal{A}_1 \) and \( 2 + \lceil K - 5 \rceil / 2 = \lceil K - 1 \rceil / 2 \) to \( \mathcal{A}_2 \). It follows that if \( n \) is odd we have \( \kappa_2^{\text{odd}} = \lceil n - 1 \rceil / 2 > n \lceil n - 1 \rceil / 2 = \kappa_1^{\text{even}} \), while if \( n \) is even we have \( \kappa_1^{\text{odd}} = n \lceil n - 1 \rceil / 2 > n \lceil n - 1 \rceil / 2 = \kappa_2^{\text{even}} \). This in turn implies that the model can contain up to \( \kappa_1^{\text{odd}} \) or \( \kappa_2^{\text{even}} \) types, respectively, consistent with generic identification.\(^{11}\)

The leave-one-out variation featured in Example 4 imposes a very specific structure on the collection \( \mathcal{A} \). We can allow more flexibility, while still achieving useful bounds on the number of identifiable types, by assuming that the menu cardinalities present in \( \mathcal{A} \) do not decline too quickly.

\(^{10}\)Observe that the even/odd partition used in Proposition 1 performs suboptimally here, producing a bound of only \( |A_2| \cdot |A_4| = 4^2 = 16 < 20 = J(\mathcal{A}) \). In fact, the even/odd partition can at best match the performance of the greedy partition, depending on the structure of the collection \( \mathcal{A} \). (A proof of this claim is available upon request.)

\(^{11}\)Since both \( \kappa_1^{\text{odd}} \) and \( \kappa_2^{\text{even}} \) exceed \( n \lceil n - 1 \rceil / 2 \), the latter bound is valid for any \( n \).
Definition 3. The menu collection \( A \) is said to be subquadratic if for each \( 1 \leq k < K - 1 \) we have \( |A_k| \leq |A_{k+1}|^2 \).

Proposition 2. Let \( A \) be subquadratic. Then \( J \leq |A_{K-1}|^{-1/2} J^*(A) \) only if the model is generically identified.

Proof. Recall the sets \( B_1(k) \) and \( B_2(k) \), for \( 1 \leq k \leq K - 1 \), constructed in the definition of the greedy partition. Let \( b_1(k) = \prod_{B_i \in B_i(k)} |B_1| \) and \( b_2(k) = \prod_{B_i \in B_i(k)} |B_2| \). To prove the result, we will show that the greedy partition of \( A \) has \( \min \{ \kappa_1, \kappa_2 \} \geq |A_{K-1}|^{-1/2} J^*(A) \).

Define the product-of-cardinalities ratio
\[
R(k) = \frac{\max \{ b_1(k), b_2(k) \}}{\min \{ b_1(k), b_2(k) \}} \geq 1.
\]  

(30)

We now 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 \( R(k + 1) = \frac{R(k)}{|A_{k+1}|} \leq |A_{k+1}| \), since \( R(k) \geq 1 \) by construction. On the other hand, if \( |A_{k+1}| < R(k) \leq |A_k| \) then
\[
R(k + 1) = \frac{R(k)}{|A_{k+1}|} \leq \frac{|A_k|}{|A_{k+1}|} \leq |A_{k+1}|, 
\]  

(31)

where the last inequality holds because \( A \) 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
\[
|A_{K-1}| \cdot \min \{ b_1(K - 1), b_2(K - 1) \} \geq \max \{ b_1(K - 1), b_2(K - 1) \}, 
\]  

(32)

\[
|A_{K-1}| \cdot \min \{ \kappa_1, \kappa_2 \} \geq \max \{ \kappa_1, \kappa_2 \}, 
\]  

(33)

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

(34)

\[
\min \{ \kappa_1, \kappa_2 \} \geq |A_{K-1}|^{-1/2} J^*(A), 
\]  

(35)

as desired. \(\square\)

Since \( J^*(A) = \prod_{k=1}^{K-1} |A_k|^{1/2} \geq |A_{K-1}|^{(K-1)/2} \), we can also conclude the following.
Corollary 1. For subquadratic $\mathcal{A}$ we have $J \leq |A_{K-1}|^{K/2-1}$ only if the model is generically identified.

Proposition 2 allows us to bound the identification capacity of a mixture choice model relative to the theoretical optimum $J^*(\mathcal{A})$. For instance, if the collection $\mathcal{A}$ is subquadratic and $|A_{K-1}| = 2$, then we know that at least a fraction $2^{-1/2} \approx 0.707$ of $J^*(\mathcal{A})$ is attainable by the greedy partition. Our next example illustrates how such facts can be used to ensure (generic) identification in an experimental design application.

Example 5. Suppose that in the experimental setting of interest, the types are ordinal (strict) preference rankings of six alternatives. With $n = 6$ there are 57 menus in total, and it may be inadvisable to require subjects to face this many choice problems. Here we wish to construct a modest collection of menus that we can be certain is powerful enough to guarantee identification in the context of $J = |\Theta| = n! = 720$ possible preference types.

By Corollary 1 it suffices (for subquadratic $\mathcal{A}$) to have $720 \leq |A_{K-1}|^{K/2-1}$; or, equivalently $K \geq 2 + 2 \log 720 / \log |A_{K-1}|$. Thus we can specify any 21 menus, or any 14 of size at least three, or any 12 of size at least four; provided the resulting collection $\mathcal{A}$ is subquadratic.

Using the full strength of Proposition 2, we can further lower the data requirements for identification: Any collection $\mathcal{A}$ of 12 menus with $\langle |A_k| \rangle_{k=1}^{12} = \langle 5, 5, 4, 4, 4, 3, 3, 3, 3, 2, 2 \rangle$ will have $J^*(\mathcal{A}) = 720 \cdot 2^{1/2}$ and hence $J = 720 = |A_{K-1}|^{-1/2} J^*(\mathcal{A})$, as required.12

12 Furthermore, returning to the original sufficient conditions in Lemma 1 allows us to drop one menu of size two, since in its absence we can still achieve $\kappa_1 = \kappa_2 = 720$ and $\kappa_3 = 2$.

Either for theoretical or for practical reasons, the researcher may need to work with low-cardinality menus (e.g., if the model is couched in terms of preference comparisons, equivalent to binary choices). Given $K$, reducing menu cardinalities of course decreases the identification power of the dataset, according to Proposition 2. But this decrease can be compensated for by a higher $K$, facilitated by a larger $n$, allowing us still to develop useful sufficient conditions. Our next two results show this concretely for the case of types as ordinal preference rankings, providing conditions for identification with low-cardinality menus and reasonable numbers of alternatives.

Corollary 2. Let $J = n!$ and $n \geq 11$. If $\mathcal{A}$ contains all possible binary menus, then the model is generically identified.
Corollary 3. Let $J = n!$ and $n \geq 4$. If $\mathcal{A}$ contains all possible binary and ternary menus, then the model is generically identified.

Indeed, the inequality $2 \log[2n!] \leq \left(\frac{n}{2}\right) \log 2$ is readily confirmed for $n \geq 11$, and similarly we have $2 \log[2n!] \leq \left(\frac{n}{2}\right) \log 2 + \left(\frac{n}{3}\right) \log 3$ for $n \geq 4$.

### 3.3 Achieving asymptotic optimality with data from all menus

Finally, we consider the prospects for identification when our dataset includes the full collection of menus. In such situations $K = 2^n - n - 1$ and the list of cardinalities is

$$\langle |A_1|, \ldots, |A_K| \rangle = \langle n, n-1, \ldots, n-1, 2, \ldots, 2 \rangle^{n \text{ times}}\langle 3 \rangle \text{ times}. \tag{36}$$

For instance, when $n = 4$ we have the cardinalities $\langle 4, 3, 3, 3, 2, 2, 2, 2, 2, 2 \rangle$; the greedy partition has $\kappa_1 = |A_1| \cdot |A_4| \cdot |A_6| \cdot |A_7| \cdot |A_9| = 96$, $\kappa_2 = |A_2| \cdot |A_3| \cdot |A_5| \cdot |A_8| \cdot |A_{10}| = 108$, and $\kappa_3 = |A_{11}| = 2$; and no value closer to the optimum $J^*(\mathcal{A}) = [10368]^{1/2} \approx 101.8$ is achievable.\(^{13}\) We record this conclusion as follows.

Proposition 3. If $n = 4$, $\mathcal{A}$ is the full collection of menus, and $J \leq 96$, then the model is generically identified.

When $n \geq 5$ there are $n \choose m$ menus of each size $m$, with a single binary menu allocated to $\mathcal{A}_3$, and so the theoretical optimum is

$$J^*(\mathcal{A}) = \left[ \frac{1}{2} \prod_{m=2}^{n} m(n \choose m) \right]^{1/2}. \tag{37}$$

Here we are able to strengthen Proposition 2 by exhibiting a lower bound on $J(\mathcal{A})$ that is asymptotically optimal. Beginning with the greedy partition, this bound is obtained by judiciously exchanging menus between the elements $\mathcal{A}_1$ and $\mathcal{A}_2$ in order to force the product-of-cardinalities ratio as close as possible to unity. We achieve a ratio of at most

\(^{13}\)Note that the even/odd partition used in Proposition 1 has $\min\{\kappa_1, \kappa_2\} = \min\{144, 72\} = 72$ in this case, and thus performs worse than the greedy partition.
and thus conclude that for practical purposes mixture choice data can distinguish up to the theoretical optimum number of types when $n$ is sufficiently large.

**Proposition 4.** If $n \geq 5$, $\mathcal{A}$ is the full collection of menus, and $J \leq \lceil \frac{n-2}{n-1} \rceil J^*(\mathcal{A})$, then the model is generically identified.

**Proof.** 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

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

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 33. If also max\{\(\kappa_1, \kappa_2\)\}/min\{\(\kappa_1, \kappa_2\)\} > \(\lceil \frac{n-1}{n-2} \rceil\), then let $m$ be the unique integer such that both $3 \leq m \leq n - 1$ and

$$\frac{m - 1}{m - 2} \leq \frac{\max\{\kappa_1, \kappa_2\}}{\min\{\kappa_1, \kappa_2\}} \leq \frac{m - 1}{m - 2}. \tag{39}$$

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 $\kappa_1'$ and $\kappa_2'$ satisfy

$$\frac{m - 1}{m} < \frac{\kappa_1'}{\kappa_2'} = \frac{\kappa_1}{\kappa_2} \left[ \frac{m - 1}{m} \right]^2 \leq \frac{[m - 1]^3}{m^2[m - 2]} < \frac{m}{m - 1}, \tag{40}$$

where the equality holds by construction, the first and second inequalities follow from Equation 39, 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

\[
\frac{m'}{m' - 1} < \frac{\max\{\kappa'_1, \kappa'_2\}}{\min\{\kappa'_1, \kappa'_2\}} \leq \frac{m' - 1}{m' - 2}.
\] (41)

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 \( \mathcal{A}_1', \mathcal{A}_2', \) and \( \mathcal{A}_3'; \) and with corresponding \( \kappa_1', \kappa_2' \) that satisfy \( \max\{\kappa_1', \kappa_2'\}/\min\{\kappa_1', \kappa_2'\} \leq [n - 1]/[n - 2] \). We then have

\[
[n - 1]\min\{\kappa_1', \kappa_2'\} \geq [n - 2]\max\{\kappa_1', \kappa_2'\},
\] (42)

\[
[n - 1][\min\{\kappa_1', \kappa_2'\}]^2 \geq [n - 2][\kappa_1'\kappa_2'] = [n - 2][J^*(\mathcal{A})]^2,
\] (43)

\[
\min\{\kappa_1', \kappa_2'\} \geq \left[\frac{n - 2}{n - 1}\right]^{1/2} J^*(\mathcal{A});
\] (44)

and the result follows.

Proposition 4 tells us that models with a very large number of types are in principle identifiable given mixture choice data from the full collection of menus.\(^\text{14}\) Depending on the application, this flexibility could be important for different reasons. For example, it allows us to more than comfortably handle the factorial growth of the number of strict preference rankings of \( n \) discrete alternatives. On the other hand, where preferences are captured by one or more parameters that are naturally continuous—such as marginal rates of substitution, discount factors, or coefficients of risk aversion—we are able to use a fine grid of permitted values for approximation in the relevant space, even in multidimensional settings. And finally, we remain free to use \( \theta \) to represent some characteristic of the agent entirely unrelated to preferences, and cannot be certain in advance that the number of types needed for an effective model will be small.

\(^{14}\)The bound on \( J \) supplied in Proposition 4 is on the order of \( 10^5 \) when \( n = 5 \), and \( 10^{13} \) when \( n = 6 \).
4 Specialized models of cognition

4.1 Deep parameter identification

The results in Section 3 ensure that the parameters \( \Omega = \langle \pi(\theta_j), \rho^\theta \rangle_{j=1}^J \) of the model in Equation 3 (namely, the distributions of types and of type-conditional choices) can be generically identified from mixture choice datasets. In applications of this methodology, however, the ultimate objects of interest will often be the “deep” cognitive parameters that impact decision making conditional on the agent’s preferences. Our framework embeds these parameters in the type-conditional SCFs \( \rho^\theta \), from which we must be able to deduce them if the specialized model is to be identified.

Section 4 addresses this deeper level of identification. We examine several well known models of bounded rationality, in each case seeking conditions under which knowledge of the type-conditional SCFs will enable us to recover the relevant cognitive parameters. Moreover, the interaction of these parameters with the agent’s preferences will typically allow us to assign an economic interpretation to each type \( \theta \), rendering moot the issue of label-swapping present in the abstract mixture choice model.

4.2 Stochastic consideration sets

We examine first the limited attention model proposed by Manzini and Mariotti [23]. In its original formulation this model assumed deterministic preferences over the alternatives, and adapting it to accommodate taste heterogeneity is desirable for applied purposes. The same is true of other recent limited attention models (such as those studied by Brady and Rehbeck [10], Cattaneo et al. [11], and Horan [16]), which can be extended to the case of heterogeneous preferences using methods similar to those outlined below.

In Manzini and Mariotti [23] the agent actively considers only a subset of the menu, the members of which are realized independently, and standard preferences are maximized over this “consideration set.” If no alternatives are considered then a default outcome \( d \notin X \) occurs, and we define both \( X^d = X \cup \{d\} \) and \( A^d = \{A \cup \{d\} : A \in A\} \). Note that the default consequence is qualitatively different from the \( n \) alternatives in \( X \), in the sense that it is never affirmatively chosen by the decision maker.
Formally, we write $\succ^\theta$ for the preference order of type $\theta$, and denote by $\gamma_y^\theta \in (0, 1)$ the probability that this type considers $y \in X$. The resulting type-conditional stochastic choice function $\rho^\theta : X^d \times A^d \rightarrow [0, 1]$ is then given by

$$
\rho^\theta(x, A) = \gamma_x^\theta \prod_{y \in A : y \succ^\theta x} [1 - \gamma_y^\theta],
$$

$$
\rho^\theta(d, A) = \prod_{y \in A} [1 - \gamma_y^\theta].
$$

In short, alternative $x$ is chosen from menu $A$ if it is considered and everything superior to it is not considered, while $d$ occurs if nothing is considered. Observe that the “salience” parameters $\gamma_y^\theta$ are independent of the menu but may depend on the agent’s type, allowing arbitrary correlation between preference and attention characteristics.

Manzini and Mariotti [23, pp. 1156–57] established that when $\Theta = \{\theta_1\}$ and $A$ is the full collection of menus, the SCF in Equations 45–46 identifies both the (deterministic) preference order $\succ^{\theta_1}$ and the $n$ salience parameters in the vector $\gamma^{\theta_1} = (\gamma_y^{\theta_1})_{y \in X}$. We can extend this finding to the present (multiple-type) setting by confirming that the type-conditional SCFs identify the $|\Theta|$ preference orders and the $|\Theta| \times n$ salience parameters in the vectors $(\gamma^\theta)_{\theta \in \Theta}$.

**Proposition 5.** If $A$ contains all possible binary menus, then both the preference orders and the salience parameters of the stochastic consideration set model are identified by the type-conditional SCFs.

**Proof.** Equation 46 implies that for each $\theta \in \Theta$ and $x, y, z \in X$ we have

$$
\frac{\rho^\theta(d, xy)\rho^\theta(d, xz)}{\rho^\theta(d, yz)} = \frac{[1 - \gamma_{x}^\theta]^2 [1 - \gamma_{y}^\theta][1 - \gamma_{z}^\theta]}{[1 - \gamma_{y}^\theta][1 - \gamma_{z}^\theta]} = [1 - \gamma_{x}^\theta]^2,
$$

and hence

$$
\gamma_{x}^\theta = 1 - \left[ \frac{\rho^\theta(d, xy)\rho^\theta(d, xz)}{\rho^\theta(d, yz)} \right]^{1/2}.
$$

Using these salience parameters, we can infer the type-specific preference orders uniquely from the fact that $x \succ^\theta y$ if and only if $\rho^\theta(y, xy) < \gamma_y^\theta$.

**Example 6.** Let the collection $A$ consist of the binary menus drawn from $X = xyz$. For
\( J = 2 \), suppose that the observed MCF leads to the type-conditional SCFs given by

\[
\rho_{\theta_1}(x, xy) = \frac{35}{50} > \rho_{\theta_1}(y, xy) = \frac{12}{50} > \rho_{\theta_1}(d, xy) = \frac{3}{50}, \quad (49)
\]
\[
\rho_{\theta_1}(x, xz) = \frac{35}{50} > \rho_{\theta_1}(z, xz) = \frac{9}{50} > \rho_{\theta_1}(d, xz) = \frac{3}{50}, \quad (50)
\]
\[
\rho_{\theta_1}(y, yz) = \frac{40}{50} > \rho_{\theta_1}(z, yz) = \frac{6}{50} > \rho_{\theta_1}(d, yz) = \frac{4}{50}, \quad (51)
\]
\[
\rho_{\theta_2}(x, xy) = \frac{27}{50} > \rho_{\theta_2}(y, xy) = \frac{20}{50} > \rho_{\theta_2}(d, xy) = \frac{3}{50}, \quad (52)
\]
\[
\rho_{\theta_2}(x, xz) = \frac{36}{50} > \rho_{\theta_2}(z, xz) = \frac{10}{50} > \rho_{\theta_2}(d, xz) = \frac{4}{50}, \quad (53)
\]
\[
\rho_{\theta_2}(d, yz) = \frac{24}{50} > \rho_{\theta_2}(y, yz) = \frac{16}{50} > \rho_{\theta_2}(z, yz) = \frac{10}{50}. \quad (54)
\]

From Equation 48 we have

\[
\gamma_{\theta_1}^x = 1 - \left[ \frac{3/50 \times 6/50}{4/50} \right]^{1/2} = 1 - 3/10 = 7/10; \quad (55)
\]

and similarly \( \gamma_{\theta_1}^y = 8/10, \gamma_{\theta_1}^z = 6/10, \gamma_{\theta_2}^x = 9/10, \gamma_{\theta_2}^y = 4/10, \) and \( \gamma_{\theta_2}^z = 2/10. \) Moreover, since \( \rho_{\theta_1}(y, xy) = 12/50 < 8/10 = \gamma_{\theta_1}^y \) we have \( x \succ_{\theta_1} y, \) and likewise \( y \succ_{\theta_1} z \succ_{\theta_2} y \succ_{\theta_2} x. \)

Observe that for type \( \theta_1, \) preference dominates attention in the sense that the preferred alternative is selected more often in each choice problem. In contrast, for type \( \theta_2 \) attention dominates preference in the sense that the more salient alternative is always chosen more frequently. \( \|
\]

One potential drawback of the proof of Proposition 5 is that it relies on the share of agents accepting the default outcome, which may not be observable in datasets used for estimation. This leads us to introduce a conditional version of the stochastic consideration set model, whose identification properties have previously been studied by Horan [16], and which we will use for the empirical application in Section 5. In this variant of the model, the SCF of preference type \( \theta \) is

\[
\rho^\theta(x, A) = \frac{\prod_{y \in A : y \succ_x} [1 - \gamma_{\theta_y}^x]}{1 - \prod_{y \in A} [1 - \gamma_{\theta_y}^x].} \quad (56)
\]

Continuing to assume that \( A \) is the full collection of menus, for \( x, y, z \in X \) we can
define the odds ratio
\[ \text{OR}_z^\theta(x, y) = \frac{\rho^\theta(x, x\bar{y}z)}{\rho^\theta(y, x\bar{y}z)} \times \frac{\rho^\theta(y, xy)}{\rho^\theta(x, xy)}. \] (57)

The type-specific preference orders are then revealed by \( x \succ^\theta y \) if and only if
\[ \max_{z \in X} \max \{ \text{OR}_z^\theta(x, y), \text{OR}_y^\theta(x, z), \text{OR}_z^\theta(z, y) \} > 1. \] (58)

Now denote by \( x_{\text{max}}^\theta \) and \( x_{\text{min}}^\theta \) the options preferred most and least (respectively) according to the inferred relation \( \succ^\theta \), and take any third option \( y \). The salience parameter for the non-extreme alternative \( y \) is revealed to be
\[ \gamma_y^\theta = 1 - [\text{OR}_{y}^\theta(x_{\text{max}}^\theta, x_{\text{min}}^\theta)]^{-1}, \] (59)
and we can compute the two remaining salience parameters as
\[ \gamma_{x_{\text{max}}^\theta}^\theta = \frac{\gamma_y^\theta \rho^\theta(x_{\text{max}}^\theta, x_{\text{max}}^\theta)}{\rho^\theta(y, x_{\text{max}}^\theta) + \gamma_y^\theta \rho^\theta(x_{\text{max}}^\theta, x_{\text{max}}^\theta)}, \] (60)
\[ \gamma_{x_{\text{min}}^\theta}^\theta = \frac{\gamma_y^\theta \rho^\theta(x_{\text{min}}^\theta, x_{\text{min}}^\theta)}{\rho^\theta(y, x_{\text{min}}^\theta) \times \frac{\gamma_y^\theta}{1 - \gamma_y^\theta}}. \] (61)

### 4.3 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, first let type \( \theta \) have preference order \( \succ^\theta \), as before, and write \( \succeq^\theta \) for the union of this order and the equality relation. Each \( \tilde{x} \in A \) may potentially be realized as the threshold alternative for menu \( A \), and this occurs with (type-dependent) probability \( \tau_A^\theta(\tilde{x}) > 0 \). 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. This implementation of satisficing (cf. Dardanoni et al. [12, Appendix B]) leads to the stochastic choice function given by
\[ \rho^\theta(x, A) = \sum_{\tilde{x} \in A : x \succeq^\theta \tilde{x}} \frac{x_A^\theta(\tilde{x})}{|\{y \in A : y \succeq^\theta \tilde{x}\}|}. \] (62)
In the satisficing context, we need each SCF $\rho^\theta$ to identify the corresponding preference order $\succ^\theta$ and threshold distributions $\langle \tau^\theta_A \rangle_{A \in \mathcal{A}}$. Once again we find that this is ensured as long as all binary menus are present in our dataset.

**Proposition 6.** If $\mathcal{A}$ contains all possible binary menus, then both the preference orders and the threshold distributions of the random satisficing threshold model are revealed by the type-conditional SCFs.

**Proof.** For each $\theta \in \Theta$, enumerate $A \in \mathcal{A}$ as $\langle z^\theta_A(1), z^\theta_A(2), \ldots, z^\theta_A(|A|) \rangle$ in such a way that $z^\theta_A(1) \succ^\theta z^\theta_A(2) \succ^\theta \cdots \succ^\theta z^\theta_A(|A|)$. Given $m < |A|$, Equation 62 then yields

$$
\rho^\theta(z^\theta_A(m), A) - \rho^\theta(z^\theta_A(m + 1), A) = \frac{\tau^\theta_A(z^\theta_A(m))}{m} > 0.
$$

(63)

This shows that $x \succ^\theta y$ if and only if $\rho^\theta(x, xy) > \rho^\theta(y, xy)$, revealing the preference types. Moreover, we can rearrange Equation 63 to write

$$
\tau^\theta_A(z^\theta_A(m)) = m \times [\rho^\theta(z^\theta_A(m), A) - \rho^\theta(z^\theta_A(m + 1), A)],
$$

(64)

and from Equation 62 we have

$$
\tau^\theta_A(z^\theta_A(|A|)) = |A| \times \rho^\theta(z^\theta_A(|A|), A).
$$

(65)

Equations 64–65 use $\succ^\theta$ and $\rho^\theta$ to reveal the threshold distribution $\tau^\theta_A$, as desired. \hfill \square

**Example 7.** Let the collection $\mathcal{A}$ consist of the binary menus drawn from $X = xyz$. For $J = 2$, suppose that mixture choice data lead to the type-conditional SCFs given by

$$
\rho^{b_1}(x, xy) = 4/5 > \rho^{b_1}(y, xy) = 1/5,
$$

(66)

$$
\rho^{b_1}(x, xz) = 9/10 > \rho^{b_1}(z, xz) = 1/10,
$$

(67)

$$
\rho^{b_1}(y, yz) = 3/4 > \rho^{b_1}(z, yz) = 1/4;
$$

(68)

$$
\rho^{b_2}(y, xy) = 9/10 > \rho^{b_2}(x, xy) = 1/10,
$$

(69)

$$
\rho^{b_2}(z, xz) = 19/20 > \rho^{b_2}(x, xz) = 1/20,
$$

(70)

$$
\rho^{b_2}(z, yz) = 3/4 > \rho^{b_2}(y, yz) = 1/4.
$$

(71)
Here the preferences are $x \succ^\theta_1 y \succ^\theta_1 z$ and $z \succ^\theta_2 y \succ^\theta_2 x$. Moreover, from Equation 64 we have $\tau^\theta_1^{xy}(x) = 4/5 - 1/5 = 3/5$, and likewise $\tau^\theta_1^{xy}(y) = 4/5$, $\tau^\theta_1^{yz}(y) = 1/2$, and $\tau^\theta_2^{xz}(z) = 9/10$.

Note that the probability $\tau^\theta_1^{xy}(x) = 3/5$ of type $\theta_1$ deliberately choosing alternative $x$ from menu $xy$ has a potential interpretation as the “intensity” of the preference $x \succ^\theta_1 y$.\textsuperscript{15} This perspective on random satisficing thresholds is pursued further by Tyson [32].

For simplicity, Proposition 6 is stated in terms of the availability of binary menus, but its proof is valid under weaker conditions. As long as any two alternatives $x$ and $y$ are present together on some menu $A \in \mathcal{A}$, Equation 63 establishes that $x \succ^\theta y$ if and only if $\rho^\theta(x, A) > \rho^\theta(y, A)$, while Equations 64–65 continue to reveal the threshold distributions.

### 4.4 Multinomial logit with menu-dependent rationality

The stochastic choice function in Equation 4 can be generalized to yield the multinomial logit specification

$$\rho^\theta(x, A) = \frac{\exp[\lambda^\theta_A u^\theta(x)]}{\sum_{y \in A} \exp[\lambda^\theta_A u^\theta(y)]}. \quad (72)$$

Here $u^\theta$ is the utility function of type $\theta$, as before, and the parameter $\lambda^\theta_A > 0$ captures this type’s degree of rationality in choosing from $A \in \mathcal{A}$. Indeed, when $\lambda^\theta_A$ is close to zero the choice of type $\theta$ from menu $A$ is essentially random, while as $\lambda^\theta_A \to \infty$ this choice approaches deterministic utility maximization. Since the functional form in Equation 72 allows the degree of rationality to depend on the menu, this framework can accommodate, for instance, preference maximization becoming more difficult when the agent faces a wider range of options. Type-dependence of the rationality parameter provides more flexibility by allowing correlation between preferences and cognition, as in the preceding models.

Given $\theta \in \Theta$, Equation 72 is invariant to the transformation $u^\theta \mapsto u^\theta + v$ for $v \in \mathbb{R}$, and accordingly we can normalize $\min_{y \in X} u^\theta(y) = 0$. Similarly, the multinomial logit SCF is invariant to $\langle \lambda^\theta_A, u^\theta \rangle \mapsto \langle \alpha \lambda^\theta_A, u^\theta / \alpha \rangle$ for $\alpha > 0$. Assuming for convenience that $X \in \mathcal{A}$, we can therefore normalize each $\lambda^\theta_X = 1$.

\textsuperscript{15}The same choice (of $x$ from $xy$) will be made randomly with probability $\tau^\theta_1^{xy}(y)/2$, yielding the total probability $\rho^\theta_1(x, xy) = \tau^\theta_1^{xy}(x) + \tau^\theta_1^{xy}(y)/2 = 3/5 + 1/5 = 4/5$ shown in Equation 66.
Under the multinomial logit model, we require the type-conditional SCFs to determine each utility function \( u^\theta \) together with the rationality parameters \( \langle \lambda_A^\theta \rangle_{A \in \mathcal{A}} \). Our next result provides sufficient conditions for identification of these model components.

**Proposition 7.** If \( \mathcal{A} \) contains all possible binary menus and the menu \( X \), then both the utility functions and the rationality parameters of the multinomial logit model are revealed by the type-conditional SCFs.

**Proof.** For \( \theta \in \Theta \) and \( x, y \in X \) we have \( u^\theta(x) > u^\theta(y) \) if and only if \( \rho^\theta(x, xy) > \rho^\theta(y, xy) \), revealing this type’s ordinal preferences. We denote by \( x_{\text{min}}^\theta \) the least preferred alternative according to these preferences (i.e., the unique option chosen less than half of the time in each binary problem). Since \( u^\theta(x_{\text{min}}^\theta) = \min_{y \in X} u^\theta(y) = 0 \), for each \( x \in X \) it follows from Equation 72 that

\[
u^\theta(x) = u^\theta(x) - u^\theta(x_{\text{min}}^\theta) = \log \rho^\theta(x, X) - \log \rho^\theta(x_{\text{min}}^\theta, X),
\]

revealing the utility function of type \( \theta \). Finally, given \( A \in \mathcal{A} \) and taking any two options \( x, y \in A \) with \( u^\theta(x) > u^\theta(y) \), Equation 72 implies that

\[
\lambda_A^\theta = \frac{\lambda_A^\theta[u^\theta(x) - u^\theta(y)]}{\lambda_X^\theta[u^\theta(x) - u^\theta(y)]} = \frac{\log \rho^\theta(x, A) - \log \rho^\theta(y, A)}{\log \rho^\theta(x, X) - \log \rho^\theta(y, X)},
\]

revealing the rationality parameters. \( \square \)

**Example 8.** Let \( X = xyz \) and let \( \mathcal{A} \) be the full collection of menus. For \( J = 2 \), consider
the SCFs given by

\[ \rho_\theta^1(x, xy) = \frac{2}{3} > \rho_\theta^1(y, xy) = \frac{1}{3}, \] (75)

\[ \rho_\theta^1(x, xz) = \frac{16}{17} > \rho_\theta^1(z, xz) = \frac{1}{17}, \] (76)

\[ \rho_\theta^1(y, yz) = \frac{8}{9} > \rho_\theta^1(z, yz) = \frac{1}{9}, \] (77)

\[ \rho_\theta^1(x, xyz) = \frac{4}{7} > \rho_\theta^1(y, xyz) = \frac{2}{7} > \rho_\theta^1(z, xyz) = \frac{1}{7}; \] (78)

\[ \rho_\theta^2(y, xy) = \frac{9}{13} > \rho_\theta^2(x, xy) = \frac{4}{13}, \] (79)

\[ \rho_\theta^2(x, xz) = \frac{4}{5} > \rho_\theta^2(z, xz) = \frac{1}{5}, \] (80)

\[ \rho_\theta^2(y, yz) = \frac{3}{4} > \rho_\theta^2(z, yz) = \frac{1}{4}, \] (81)

\[ \rho_\theta^2(y, xyz) = \frac{3}{6} > \rho_\theta^2(x, xyz) = \frac{2}{6} > \rho_\theta^2(z, xyz) = \frac{1}{6}. \] (82)

Here the preferences are \( x \succ_\theta^1 y \succ_\theta^1 z = x_{\text{min}} \) and \( y \succ_\theta^2 x \succ_\theta^2 z = x_{\text{min}} \). Moreover, from Equation 73 we have \( u_\theta^1(x) = \log \frac{4}{7} - \log \frac{1}{7} = \log 4 \), and likewise \( u_\theta^2(y) = \log 3 \) and \( u_\theta^1(y) = u_\theta^2(x) = \log 2 \). From Equation 74 we have also

\[ \lambda_{xy}^\theta = \frac{\log \frac{9}{13} - \log \frac{4}{13}}{\log \frac{3}{6} - \log \frac{2}{6}} = 2, \] (83)

and similarly \( \lambda_{xy}^{\theta_1} = \lambda_{yz}^{\theta_2} = 1, \lambda_{xz}^{\theta_1} = \lambda_{xz}^{\theta_2} = 2, \) and \( \lambda_{yz}^{\theta_1} = 3 \).

### 4.5 Fechnerian models

The stochastic choice function \( \rho^\theta \) is said to be Fechnerian (cf. Strzalecki [29]) if it admits strictly increasing functions \( \varphi_m^\theta : \mathbb{R}^{m-1} \rightarrow [0, 1] \) such that for each \( x \in A \) we have

\[ \rho^\theta(x, A) = \varphi_{|A|}^\theta(\lambda_A^\theta(u^\theta(x) - u^\theta(y))_{x \neq y \in A}). \] (84)

In other words, \( \rho^\theta(x, A) \) is determined by the relevant utility differences \( u^\theta(x) - u^\theta(y) \), scaled by a (possibly menu-dependent) sensitivity parameter \( \lambda_A^\theta > 0 \). The multinomial logit model is Fechnerian in this sense, with binary choice probabilities determined by \( \varphi_2^\theta(v) = [1 + \exp[-v]]^{-1} \) and analogous transformations applying to larger menus. The

\[ 16 \text{Models of this sort were originally proposed by Fechner [14] and Thurstone [30].} \]
same property is exhibited by a version of the random threshold model (see Tyson [32]), with associated binary transformation

\[
\phi_2^\theta(v) = \begin{cases} 
1 - [1/2] \exp[-v] & \text{if } v \geq 0, \\
[1/2] \exp v & \text{if } v < 0.
\end{cases}
\] (85)

In contrast, the stochastic consideration set model is fundamentally non-Fechnerian, since for \(x \succ^\theta y\) the probability \(\rho^\theta(x,xy) = \gamma^\theta_x\) is independent of \(y\) and thus cannot vary with \(\lambda^\theta_{xy}[u^\theta(x) - u^\theta(y)]\).

Now let \(A\) contain all possible binary menus and consider the general specification in Equation 84. Here \(\rho^\theta(x,A)\) is observed (or inferred from MCF data), the transformations \(\phi^\theta_t\) are known, and both the utility function \(u^\theta\) and the sensitivity parameters \(\lambda^\theta_A\) are unknown. Since \(\phi^\theta_2\) is strictly increasing and \(\lambda^\theta_{xy} > 0\), we have that \(u^\theta(x) > u^\theta(y)\) if and only if

\[
\rho^\theta(x,xy) = \phi^\theta_2(\lambda^\theta_{xy}[u^\theta(x) - u^\theta(y)]) > \phi^\theta_2(\lambda^\theta_{xy}[u^\theta(y) - u^\theta(x)]) = \rho^\theta(y,xy), \tag{86}
\]

revealing the ordinal preferences of type \(\theta\). Our goal is then to use these preferences, our knowledge of the \(\phi^\theta_t\) functions, and any appropriate ancillary assumptions to identify the cardinal utilities and sensitivity parameters.

To illustrate how this analysis could proceed further, suppose that each \(\lambda^\theta_{xy} = \lambda^\theta_2 > 0\); that is, the sensitivity parameter is constant over the collection of binary menus. Denoting the worst option by \(x^\theta_{\min}\), as before, we can normalize \(u^\theta(x^\theta_{\min}) = 0\) and \(\lambda^\theta_2 = 1\) to find \(u^\theta(y) = [\phi^\theta_2]^{-1}(\rho^\theta(y,x^\theta_{\min}y))\) for \(y \neq x^\theta_{\min}\). Each instance of Equation 84 with \(|A| > 2\) then has \(\lambda^\theta_A\) as the only remaining unknown, and solving for these parameters will complete the identification exercise.

In the empirical application in Section 5 we will use an ordinal analog of the above class of SCFs, to be called the “quantal” Fechnerian model. To define this model, we again enumerate \(A \in A\) as \(\langle z^\theta_A(1), z^\theta_A(2), \ldots, z^\theta_A(|A|)\rangle\) so that \(z^\theta_A(1) \succ^\theta z^\theta_A(2) \succ^\theta \cdots \succ^\theta z^\theta_A(|A|)\). For “decay” parameter \(\delta^\theta_A \in (0, 1)\), the probability of the \(m\)th ranked alternative being
chosen is then given by
\begin{equation}
\rho^\theta(z_A^\theta(m), A) = \frac{[\delta_A^\theta|A|^{-1}][1 - \delta_A^\theta]}{1 - [\delta_A^\theta]|A|}.
\end{equation}
This yields the relation 
\begin{equation}
\rho^\theta(z_A^\theta(m + 1), A) = \delta_A^\theta \times \rho^\theta(z_A^\theta(m), A),
\end{equation}
so that choice probability declines geometrically with ordinal rank position.

5 Empirical application

5.1 Model formulation

We now proceed to outline how the theoretical identification framework in Sections 2–4 can be implemented in practice, illustrating this with a “proof-of-concept” application to a preexisting experimental dataset.

Enumerating \( C \) as \( \langle c_i \rangle_{i=1}^I \), we can write the function \( \mu \) in vector form as \( \mu = \langle \mu(c_i) \rangle_{i=1}^I \), and similarly we write \( \pi = \langle \pi(\theta_j) \rangle_{j=1}^J \) for the distribution of preference types. The general mixture choice model in Equation 3 can then be expressed as \( \mu = R \pi \), where the matrix
\begin{equation}
R = \begin{bmatrix}
\prod_k \rho^{\theta_1}(c_1(A_k), A_k) & \prod_k \rho^{\theta_2}(c_1(A_k), A_k) & \cdots & \prod_k \rho^{\theta_J}(c_1(A_k), A_k) \\
\prod_k \rho^{\theta_1}(c_2(A_k), A_k) & \prod_k \rho^{\theta_2}(c_2(A_k), A_k) & \cdots & \prod_k \rho^{\theta_J}(c_2(A_k), A_k) \\
\vdots & \vdots & \ddots & \vdots \\
\prod_k \rho^{\theta_1}(c_I(A_k), A_k) & \prod_k \rho^{\theta_2}(c_I(A_k), A_k) & \cdots & \prod_k \rho^{\theta_J}(c_I(A_k), A_k)
\end{bmatrix}
\end{equation}
describes the transition from unobserved type data to observed mixture choice data.

From the present perspective, a specialized model of cognition (see Section 4) amounts to a set of restrictions on \( R \) involving the model-dependent cognitive parameters we wish to estimate. Collecting these parameters in \( \beta = \langle \beta_\ell \rangle_{\ell=1}^L \), we write \( R(\beta) \) for the associated transition matrix and \( \rho^\theta_\beta \) for the SCF conditional on preference type \( \theta_\beta \).

Consider an i.i.d. sample \( Y = \langle Y_s \rangle_{s=1}^S \) of size \( S \), where each \( Y_s \in C \) records the choice behavior of subject \( s \) over a list \( \langle A_k \rangle_{k=1}^K \) of menus. The probability of observing \( Y_s = c_i \) is computed as \( \mu(Y_s) = \mu(c_i) = \mu_i = R(\beta)\pi \), and the log-likelihood of the entire sample then appears as

29
\[ \text{LL}(Y | \pi, \beta) = \sum_{s=1}^{S} \log \left[ \sum_{j=1}^{J} \pi(\theta_j) \prod_{k=1}^{K} \rho_{\beta} (Y_s(A_k), A_k) \right]. \quad (89) \]

Direct numerical maximization of this function by choice of \( (\pi, \beta) \) is possible in theory, but may in practice be computationally challenging due to instability of the objective. We instead employ the Expectation Maximization (EM) algorithm, at each step imposing constraints appropriate to the specialized model of cognition being estimated (e.g., each \( \gamma_\theta \in [0, 1] \) for the stochastic consideration set model). \(^{17}\)

5.2 Experimental dataset

The dataset used for our proof-of-concept application is from the experiment reported in Manzini and Mariotti \cite{21}, which was designed to study time preference and not to test any of the cognitive models examined in this paper. It is relevant here primarily because choices were elicited from all menus drawn from a grand set of four alternatives, yielding MCF data appropriate for our framework. Note that with \( n = 4 \) there are \( J = n! = 24 \) ordinal preference types.

The incentivized experiment involved 102 subjects choosing from menus of monetary payments received in three installments over a period of nine months. \(^{18}\) The four payment schedules are summarized in Table 1, and may be labeled I (increasing), D (decreasing), C (constant), and J (jump). Subjects were asked to choose from each of the eleven menus arising from \( X = \{I, D, C, J\} \); of which six are binary, four ternary, and one universal.

We estimate three of the specialized models from Section 4: stochastic consideration sets (conditional on no default), random satisficing thresholds, and the quantal Fechnerian model. Under stochastic consideration there are four salience parameters for each type,\(^{17}\)

\(^{17}\)Our framework is an instance of a latent class model, where the response variables (i.e., the choices from each \( A \)) are independent conditional on the class (i.e., the type \( \theta \)). 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 \cite{25}). Here the log-likelihood expression in Equation 89 is maximized by alternating the following two steps until convergence is achieved: [1.] Given the observed choices \( Y \) and the current parameter estimates \( \hat{\beta} \), compute the expectation of log-likelihood with respect to the current conditional type distribution \( \hat{\pi} \). [2.] Update the parameter estimates in \( \hat{\beta} \) by maximizing the expected log-likelihood found in the first step.

\(^{18}\)Each subject earned a €5 participation fee and had a 50% chance of being drawn for payment, in which case one menu was selected at random and the subject’s choice implemented.
## 5.3 Estimated preference distributions

We estimate the preference type distribution and the cognitive parameters of each model using the EM algorithm (see Footnote 17). For brevity we focus here on the preference distributions for the three selected models, summarized in Figure 1. This figure displays the estimated frequency of each of the \( J = 24 \) types, enumerated as shown in Table 2.

Stochastic consideration sets, random satisficing thresholds, and the quantal Fechnerian model largely agree about the distribution of preferences in our experimental dataset. Under all three models, approximately 50% of subjects are assigned to preference order DCIJ (type \( \theta_9 \)), while roughly 15% are assigned to CDIJ (\( \theta_{15} \)) and 5% to CIJD (\( \theta_{14} \)). Options D and C are ranked above options I and J by close to 65% of the population, and very few subjects place I or J in first position. The only notable disagreement concerns DCJI (\( \theta_{10} \)), which the quantal Fechnerian model attributes to 9% of subjects as compared

<table>
<thead>
<tr>
<th>delay</th>
<th>I (increasing)</th>
<th>D (decreasing)</th>
<th>C (constant)</th>
<th>J (jump)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 months</td>
<td>€8</td>
<td>€24</td>
<td>€16</td>
<td>€8</td>
</tr>
<tr>
<td>6 months</td>
<td>€16</td>
<td>€16</td>
<td>€16</td>
<td>€8</td>
</tr>
<tr>
<td>9 months</td>
<td>€24</td>
<td>€8</td>
<td>€16</td>
<td>€32</td>
</tr>
</tbody>
</table>

Table 1: Alternative payment schedules in the time preference experiment.

<table>
<thead>
<tr>
<th>IDCJ ( \mapsto ) 1</th>
<th>IJDC ( \mapsto ) 5</th>
<th>DCIJ ( \mapsto ) 9</th>
<th>CIDJ ( \mapsto ) 13</th>
<th>CJID ( \mapsto ) 17</th>
<th>JDIC ( \mapsto ) 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDJC ( \mapsto ) 2</td>
<td>IJCD ( \mapsto ) 6</td>
<td>DCJI ( \mapsto ) 10</td>
<td>CIJD ( \mapsto ) 14</td>
<td>CJDI ( \mapsto ) 18</td>
<td>JDCI ( \mapsto ) 22</td>
</tr>
<tr>
<td>ICDJ ( \mapsto ) 3</td>
<td>DICJ ( \mapsto ) 7</td>
<td>DJIC ( \mapsto ) 11</td>
<td>CDIJ ( \mapsto ) 15</td>
<td>JIDC ( \mapsto ) 19</td>
<td>JCID ( \mapsto ) 23</td>
</tr>
<tr>
<td>ICJD ( \mapsto ) 4</td>
<td>DIJ C ( \mapsto ) 8</td>
<td>DJCI ( \mapsto ) 12</td>
<td>CDJI ( \mapsto ) 16</td>
<td>JCID ( \mapsto ) 20</td>
<td>JCDI ( \mapsto ) 24</td>
</tr>
</tbody>
</table>

Table 2: Enumeration of preference types for the purposes of Figures 1–2. (For example, we have \( I \succ^\theta_1 D \succ^\theta_1 C \succ^\theta_1 J \) and \( J \succ^\theta_{24} C \succ^\theta_{24} D \succ^\theta_{24} I \).)
As seen above, our choice dataset is modest in size ($102 \times 11$) and our models have numerous cognitive parameters (96, 144, and 264), casting doubt on the robustness of the estimation. To address this issue, we can require that the cognitive parameters are constant across preference types, leading to a “restricted” version of each model. This extra constraint reduces the cognitive parameter count to four for stochastic consideration sets, six for random satisficing thresholds, and eleven for the quantal Fechnerian model. The resulting estimated preference distributions, displayed in Figure 2, are reassuringly similar to the unrestricted case.

For both the unrestricted and restricted versions of the three selected models, Table 3 reports the values of three model selection criteria: the maximized log-likelihood (MLL), the Bayesian information criterion (BIC), and the Akaike information criterion (AIC). Although the experiment underlying our dataset was not designed for model selection, we can observe that the restricted versions of each model score uniformly better than the unrestricted versions. This suggests that allowing type-dependence leads to an unnecessary proliferation of cognitive parameters in the choice environment under investigation.

Figure 1: Estimated distributions of preference types, as enumerated in Table 2, for the three indicated models.
Figure 2: Estimated distributions of preference types, as enumerated in Table 2, for the restricted versions of the three indicated models.

<table>
<thead>
<tr>
<th>cognitive model</th>
<th>parameters</th>
<th>model selection criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>stochastic consideration sets</td>
<td>unrestricted</td>
<td>MLL: -333.88, BIC: 1222.81, AIC: 907.76</td>
</tr>
<tr>
<td></td>
<td>restricted</td>
<td>MLL: -384.30, BIC: 898.10, AIC: 824.60</td>
</tr>
<tr>
<td>random satisficing thresholds</td>
<td>unrestricted</td>
<td>MLL: -331.92, BIC: 1440.86, AIC: 999.84</td>
</tr>
<tr>
<td></td>
<td>restricted</td>
<td>MLL: -383.32, BIC: 905.38, AIC: 826.63</td>
</tr>
<tr>
<td>quantal Fechnerian</td>
<td>unrestricted</td>
<td>MLL: -291.73, BIC: 1915.53, AIC: 1159.50</td>
</tr>
<tr>
<td></td>
<td>restricted</td>
<td>MLL: -390.58, BIC: 943.05, AIC: 851.17</td>
</tr>
</tbody>
</table>

Table 3: Evaluation of six estimated models by three selection criteria. The “restricted” version of each model disallows type-dependence of cognitive parameters.
6 Concluding remarks

We view mixture choice functions as a device for representing data that is often available in practice, but has mostly been ignored by traditional choice theory. The essential property of this data is that it includes the joint distribution of the behavior of a population across a series of choice “occasions” (cf. Dardanoni et al. [12, p. 1285]). In this paper we structure occasions as the domain of a choice function, leading to the MCF formalism described in Section 2. But other interpretations of the concept are possible, and will result in models that can be studied using tools similar to those employed here.

For example, we could treat occasions as repeated choices from a menu of alternatives that remain substantially the same, but vary over time on particular dimensions such as price or location. Our dataset would then resemble a panel; for instance, a population of homeowners who regularly remortgage their properties with one of a fixed set of potential lenders. Likewise, we could interpret 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 would offer the prospect of “market research” applications such as modeling the use of scanner or loyalty-card data to infer the characteristics of the population of consumers in terms of both preference and persuasion.

References


