Identifying Heterogeneous Decision Rules From Choices When Menus Are Unobserved

Larry G. Epstein[a]
[a]Corresponding author. larry.epstein@mcgill.ca

Kaushil Patel[b]
[b]kaushil.patel@mail.mcgill.ca
[a,b]Department of Economics, McGill University
855 Sherbrooke St W., Montreal, Quebec H3A 0C4

January 5, 2026

Abstract

Consider aggregate choice data from a population with heterogeneity in both preferences (or more general decision rules) and in menus, and where the analyst has limited information about how menus are distributed across the population. We determine what can be inferred from aggregate data about the distribution of preferences by identifying the set of all distributions that are consistent with the preceding. Our main theorem strengthens and generalizes existing results on such identification and provides an alternative analytical approach (using convex capacities) to study the problem.

Keywords: discrete choice, partial identification, unobserved heterogeneity, convex capacities, core

1 Introduction

1.1 Motivation and outline

Consider the problem of explaining the distribution of choices in a heterogeneous population. Denote by λ the probability distribution of chosen alternatives, the data. A common approach is to posit heterogeneity in decision rules (or underlying preferences) and possibly also in the menus from which alternatives are chosen. A decision rule d specifies the alternative $d(A)$ chosen from each menu A; the set of all decision rules is $\mathcal{D}$. An individual with decision rule d faces menu A with probability $\pi_d(A)$. Decision rules are distributed according to a probability measure Q that is to be inferred from the data, while the collection of probability measures $\{\pi_d\}_{d\in\mathcal{D}}$ is known to the analyst (possibly up to unknown parameters).[1] Accordingly, she seeks Q satisfying, *for the given* $\{\pi_d\}_{d\in\mathcal{D}}$,

$$\lambda(a) = \sum_d \sum_A Q(d)\, \pi_d(A)\ \mathbf{1}_{d(A)=a}, \tag{1.1}$$

for all alternatives a. Then empirical frequencies are rationalized by the heterogeneity in decision rules described by Q. Of particular interest is the set of all rationalizing Qs (the sharp identified region).

The above model is general in that it covers the bulk of the discrete choice literature where various special cases are adopted;[2] for example, the traditional assumption (McFadden 1974) that the menu corresponding to each choice is observed corresponds to the special case where, for each d, $\pi_d(A_d) = 1$ for some A_d. However, data about menus that would support knowledge of the conditional probabilities π_d are often unavailable (see Manski (1977) and the overviews and many references in Barseghyan et al (2021, pp. 2016-2017, 2041-2043) and Azrieli and Rehbeck (2025)). Notably, decision models based on consideration sets (Abaluck and Adams-Prassl 2021, Cattaneo et al 2020, Manzini and Mariotti 2014, Masatlioglu et al 2012) or rational inattention (Caplin et al 2019) view choices as made from subjective menus, thus arguing against their observability.[3] One is led to the concern

[1] Filiz-Ozbay and Masatlioglu (2023) call this a random-choice model (RCM), defined by a probability distribution over a collection of choice functions (potentially irrational). They axiomatize a specific class of RCMs under the assumption of rich stochastic choice data.

[2] We are ignoring covariates that often appear in this literature, and that could be added below, because they are not germane to our contribution. We adopt a streamlined formulation in order to maximize transparency of the main point of the paper.

[3] To be clear, we use "menu" to refer to the set from which an alternative is chosen

that conclusions about the identified set of measures Q that are based on (1.1) sometimes rely on ad hoc assumptions about menus.

An objective of this paper is to robustify the above model by incorporating the analyst's imperfect knowledge about menus. One alternative to the perfect information assumption is complete ignorance about menus - "anything goes" for specifications of π_ds. However, in general, one would expect there to be partial information about the menu process which, if exploited, would permit sharper identification. Therefore, we admit a range of assumptions about the analyst's information that are intermediate between complete ignorance and perfect knowledge. In all cases, we show (Theorem 2.1) that the implied identified set of distributions consists of all measures Q satisfying a finite set of linear inequalities and hence forms a polytope (a convex set with finitely many extreme points); in particular, it is computationally tractable.

Our formulation uses convex (or supermodular) capacities and their cores. (The appendix collects the few basic definitions and facts regarding capacities that are used below; a very accessible and comprehensive reference is Grabisch (2016).) Capacities are set functions that generalize probability measures in order to permit a role in the representation of beliefs for limited information and the resulting limited confidence in any single probability measure - in other terms, uncertainty about probabilities, often termed ambiguity. They arise in decision theory, notably in Schmeidler's (1989) Choquet expected utility theory, where convexity of the capacity is identified with aversion to such uncertainty and where convexity characterizes the Choquet models that conform also to multiple-priors utility (Gilboa and Schmeidler 1989).[4] For our purposes, the key technical feature of convex capacities is that "*the core of a mixture of capacities equals the mixture of their cores*" (see (A.4) for a formal statement). Given our formulation, this property leads to a short transparent (indeed elementary) proof of our theorem that applies to and unifies all of our specifications. We view this simplicity and the associated epistemic perspective as a strength and a contribution.

The scope of our results merits emphasis. Thus far we have interpreted

by maximizing preference or by applying another decision rule. Consequently, it may be a strict subset of the objective feasible set, (for example, a consideration set), that is determined by the individual's cognitive deliberation process and is unobservable to the analyst.

[4]Convex capacities, or equivalently their conjugates, known as 2-alternating, are important also in statistical theory (in proving an extension of the Neyman-Pearson Lemma (Huber and Strassen 1973) and in supporting a version of Bayes' theorem for capacities (Wasserman and Kadane 1990). They appear also in cooperative game theory as characteristic functions. However, the epistemic interpretation is a better fit here.

the paper as addressing heterogeneity in choice assuming heterogeneity in decision rules and the unobservability of menus. However, with suitable reinterpretation of the symbols in the formal model, Theorem 2.1 *applies also to other contexts where one seeks the identification of a heterogeneous characteristic of prime interest that is robust to other unobservables.* A dual interpretation addresses identification of the distribution of menus given partial (or complete) ignorance about the distribution of decision rules (section 2.1). Two additional reinterpretations are: (i) Identification of the distribution of threshold levels in a population of satisficing decision makers given their choices but where individuals differ in the (unobserved) order in which they consider alternatives (section 2.4). (ii) Identification of the distribution of effort in a population of workers who share common observable characteristics (e.g. education and experience) and who work independently, given the empirical frequency distribution of outputs but where other factors that may influence output are poorly understood (section 2.5).

1.2 Related literature

First we relate our contribution to some recent papers in discrete choice (and related econometrics) that also weaken a priori assumptions about menus. Barseghyan et al (2021) study identification in a random utility model where the distribution of menus in the population is unknown. Two relatively minor differences from our model are that: they assume preference maximization (particularly, Sen's α condition) rather than general decision rules, and they assume that all menus of size at least κ, ($\kappa \geq 2$), a parameter specified by the analyst, are conceivable for any individual conditional on her preference order, while we allow the set of conceivable menus to be arbitrary. An example in the Supplementary Appendix illustrates that admitting some singleton menus can affect identification.) More importantly, they deal *only with the case of complete ignorance* of the menu process, for which their characterization of the sharp identified set corresponds (apart from their inclusion of covariates) to our complete-ignorance result in Theorem 2.1.[5] This difference from the present paper is reflected in a difference in proofs. Barseghyan et al (2021) applies the theory of random sets. This approach is limiting because, as is well known, each random set can be identified with a belief function which is a very special kind of convex capacity that precludes many of the richer information structures (those short of complete ignorance) that are accommodated in our theorem. Lu (2022) assumes that all conceivable menus are bounded above and below in the sense of set inclusion, and that

[5] Minor differences are described following the statement of our theorem in section 2.3.

the bounding menus are *known.* He uses the latter, and the assumption that decision rules satisfy Sen's α, to describe a superset of the identified region. In contrast, our conditions on Q are both necessary and sufficient for Q to rationalize the data, thus yielding the sharp identified region. Azrieli and Rehbeck (2025) also study what can be learned from aggregate choice frequencies, but with several differences from the present paper. A major difference is that they assume that the marginal empirical distributions of both menus and choices are known (constitute the data). In their study of random utility models, they assume that menus are homogeneous across decision makers, that is, the distribution of menus does not depend on the decision rule, while we allow for such dependence. Where menus are based on consideration one would expect them to depend on preference (or decision rule), as in the applied papers by Goeree (2008), and Abaluck and Adams-Prassl (2021). Regarding proof arguments, Azrieli and Rehbeck also highlight their use of "known properties of the core," though these do not include the key property that we exploit here, and they borrow more from cooperative game theory than from decision theory and thus do not emphasize epistemics in their interpretations. Further, their proofs (specifically for their Proposition 9) use not only core properties but also network flow arguments (based on a version of Hall's marriage theorem), while we use only the mixture property of the core noted above.

Dardanoni et al (2020) also explore what can be inferred from aggregate choice data, though their focus is on cognitive heterogeneity rather than on preference (or decision rule) heterogeneity. Individuals differ in cognitive "type" and, given an objective feasible set, they arrive at different consideration sets (menus in our terminology); further, they do so in a way that conforms to specific functional forms - the "consideration capacity model" (which limits the cardinality of the consideration set) or the "consideration probability model" (Manzini and Mariotti 2014). In the section most closely related to our paper, where preferences are unobservable and heterogeneous, they assume that choices are observed from multiple "occasions" across which both the feasible set and cognitive heterogeneity are stable. With this rich dataset and functional form restrictions they prove point identification of the distribution of cognitive types in the consideration capacity model. Roughly speaking, from the perspective of our formal framework, they severely restrict the distribution of decision rules and aim at identification of the menu formation process (π_d), which reflects the distribution of cognitive type. Unsurprisingly, their proof arguments are much different than ours.

Doval and Eilat (2023) study the setting where the analyst knows the marginal over an agent's actions and the prior over states of the world, but does not know the distribution of actions given realizations of the states of the world. They ask when two such marginals (over actions and states, similar to the dataset in Azrieli and Rehbeck (2025) can be rationalized in the sense of a Bayes correlated equilibrium as the outcome of the agent learning something about the state before taking an action. Their characterization result is two systems of linear inequalities that are necessary and sufficient for the dataset to be consistent with a Bayes correlated equilibrium. One of these can be established using our "mixture of cores" property. Their proof relies partly on network flow arguments.

Galichon and Henry (2011) are, to the best of our knowledge, the first to demonstrate the usefulness of convex capacities for characterizing partially identified sets. Their context, which differs from ours, is the identification of structural parameters in models with normal form games having multiple mixed strategy equilibria and where little is understood about selection. Another difference is that they do not use the "mixture of cores" property that is central to this paper.

2 Robust identification

2.1 Rationalizability

The (finite) universal set of alternatives is X, and the set of probability distributions or measures on X is denoted $\Delta(X)$. Each individual in a finite population faces a menu, a subset of X, from which she chooses one alternative. The collection of all "relevant" menus is denoted $\mathcal{A}$, with generic element A. The collection $\mathcal{A}$ is a primitive, determined by the analyst. (Barseghyan et al (2021) take $\mathcal{A}$ to be the set of all menus with cardinality at least 2.) Another primitive is a finite set $\mathcal{D}$ of decision rules, where, for each d in $\mathcal{D}$, $d(A)$ denotes the alternative that d chooses from the menu A in $\mathcal{A}$. We do not impose any requirements on decision rules, for example, they need not be derived from preference maximization. The data to be explained are represented by $\lambda \in \Delta(X)$, the empirical frequency distribution of chosen alternatives across the population.

The analyst's view of the menu formation process determines what constitutes an "explanation." We assume that she is certain that only menus in $\mathcal{A}$ are relevant, but otherwise she has limited understanding of how menus are determined; in particular, for every d, she cannot be confident in any single conditional probability distribution over menus $\pi_d \in \Delta(\mathcal{A})$, which suggests modeling via a *set* $\Pi_d \subset \Delta(\mathcal{A})$ of conditional distributions. *These sets are not "data", but rather are subjective, chosen by the analyst, in a way that captures her limited confidence and desired robustness* much as sets of priors are interpreted in the maxmin model of decision-making (Gilboa and Schmeidler (1989)).

Say that $Q \in \Delta(\mathcal{D})$ *rationalizes* λ *given* $\{\Pi_d\}$ *if, for all* d, *there exists* $\pi_d \in \Pi_d$ such that

$$\lambda(a) = \sum_d \sum_A Q(d)\, \pi_d(A)\; \mathbf{1}_{d(A)=a}, \ \text{ for all } a \in X. \tag{2.1}$$

Then empirical frequencies are rationalized by the heterogeneity in decision rules described by Q. Of particular interest is the set of all rationalizing Qs (the sharp identified region).

The extreme case where the analyst knows the distribution over menus, as in the standard literature centered on (1.1), is accommodated by taking Π_d to be a singleton set for every d. The other extreme of complete ignorance about how menus are distributed, (as in Barseghyan et al (2021), for example), is expressed by taking $\Pi_d = \Delta(\mathcal{A})$ for each d, whereby any distribution of menus is viewed as conceivable. The associated robustness may be desirable but comes with costs (that we formalize below). First, if "anything goes," then the identified region for any given data λ is large. Second, with such weak maintained assumptions, (almost) every λ can be rationalized by some Q (as clarified in the next subsection). Consequently, and also because there are situations in which there exists partial information about the menu process and less robustness is desired, we propose a model that also accommodates intermediate situations. The moderation of complete ignorance takes the form that, for each d, only distributions of menus in Π_d are conceivable. More fully, and what may not be as clear from the formalism, for any pair of decision rules d and d', any $\pi_d \in \Pi_d$ is possible in conjunction with any $\pi_{d'} \in \Pi_{d'}$. Also implied by the notion of rationalizability in (2.1) is that every conceivable joint distribution of menus across all decision rules is a product measure - the analyst is certain that the menu realization for d does not affect the marginal menu distribution for any other d'. This is a feature also of the benchmark complete ignorance model (Barseghyan et al (2021)), which is, after all, a special case. Further, such independence is defensible in a range of situations. For example, where menus are consideration sets determined by the subjective proclivities of individuals, there would often, or arguably even typically, be little reason to postulate that the realized consideration menu for d is informative about that for d'.

Finally, we point out that reinterpretation of the formal primitives $\mathcal{D}$ and $\mathcal{A}$ renders the notion of rationalizability and the analysis to follow applicable to additional settings where one seeks to characterize the robust identification of heterogeneity (see section 2.4 for a reinterpretation dealing with satisficing behavior). Here we outline a different example. Switch current interpretations so that each d denotes a menu, each A denotes a decision rule (A maps $\mathcal{D}$ into X), and $d(A)$ denotes the alternative chosen from menu d given the decision rule A (alternatively, one could write $A(d)$ in place of $d(A)$). Then the paper, including the main result Theorem 2.1 that characterizes rationalizability, admits a dual interpretation as addressing identification of the distribution of menus given partial (or complete) ignorance about the distribution of decision rules. We focus on the existing interpretation - the identification of the distribution of preferences (decision rules) - because of the common presumption that preferences, unlike menus, might be expected to be invariant across choice problems thus permitting predictions of choice for other settings. However, though we do not pursue it further, the dual analysis and identification of menu formation, where menus are understood as consideration sets, might be valuable along the lines indicated in the introduction (Dardanoni et al 2020, Manzini and Mariotti 2014). For example, it could help to discriminate empirically between alternative models of consideration set formation, or to identify cognitive heterogeneity (suitably defined).

2.2 Some specifications

Here we describe some particular specifications for the sets $\{\Pi_d\}$.[6]

Complete ignorance: Let $\Pi_d = \Delta(\mathcal{A})$ for each d as indicated above.

ϵ-**contamination**: For each d, let $\widehat{\pi}_d \in \Delta(\mathcal{A})$ be a focal probability distribution over menus, perhaps the analyst's best "point estimate," but one in which she may not have complete confidence. As a reflection of her incomplete confidence she entertains as possible all contaminations of $\widehat{\pi}_d$ of the form

$$\pi_d = (1-\epsilon)\,\widehat{\pi}_d + \epsilon\widetilde{\pi}_d,$$

where $\widetilde{\pi}_d$ is any measure on $\mathcal{A}$ and where $0 \leq \epsilon \leq 1$ is a parameter to be specified by the analyst. That is, let

$$\begin{aligned}\Pi_d &= \{\pi_d : \pi_d = (1-\epsilon)\,\widehat{\pi}_d + \epsilon\widetilde{\pi}_d,\ \widetilde{\pi}_d \in \Delta(\mathcal{A})\} \\ &= (1-\epsilon)\,\widehat{\pi}_d + \epsilon\Delta(\mathcal{A}).\end{aligned} \tag{2.2}$$

The extremes $\epsilon = 0, 1$ correspond respectively to the complete confidence and complete ignorance models respectively. Further, it is easy to see that

[6] They are all well-known in both robust statistics and in decision theory. We have borrowed them and their properties described later from Wasserman and Kadane (1990). However, we have not seen the last four used previously in the present context.

Π_d grows larger in the sense of set inclusion as ϵ increases in $[0, 1]$. This suggests the interpretation of decreasing confidence (or increasing ignorance) as ϵ increases.

The "ϵ-contamination" model has been used frequently in robust statistics (e.g. Huber (1964), Huber and Ronchetti 2009, Wasserman and Kadane 1990), and also in decision theory and its many applications where it is a useful parametric specialization of the set of priors appearing in multiple-priors utility (Gilboa and Schmeidler 1989).

Variation neighborhood: For any p' and p in $\Delta(\mathcal{A})$, define the distance between them by

$$\delta(p', p) = \sup_{A \in \mathcal{A}} \mid p'(A) - p(A) \mid .$$

Fix a reference/focal measure P_d on $\mathcal{A}$ and $\epsilon_d > 0$, and let

$$\Pi_d = \{p \in \Delta(\mathcal{A}) : \delta(p_d, P_d) < \epsilon_d\}. \tag{2.3}$$

Interval beliefs: Let p_{*d} and p_d^* be measures (not probability measures) on $\mathcal{A}$, satisfying

$$p_{*d}(\cdot) \leq p_d^*(\cdot) \text{ and } 0 < p_{*d}(\mathcal{A}) < 1 < p_d^*(\mathcal{A}),$$

and define

$$\Pi_d = \{p_d \in \Delta(\mathcal{A}) : \ p_{*d}(\cdot) \leq p_d(\cdot) \leq p_d^*(\cdot) \text{ on } \mathcal{A}\}.$$

In the special case

$$p_{*d} = a_d P_d \text{ and } p_d^* = b_d P_d,$$

where $a_d < 1 < b_d$ and P_d is a (fixed) probability measure on $\mathcal{A}$, one obtains

$$\Pi_d = \{p_d \in \Delta(\mathcal{A}) : \ a_d P_d \leq p_d \leq b_d P_d\}.$$

2-dimensional beliefs: Let P_d^1 and P_d^2 be two distinct probability measures on $\mathcal{A}$. The analyst views these measures and all averages (mixtures) as the set of relevant probability laws. Accordingly,

$$\Pi_d = \{\alpha P_d^1 + (1-\alpha) P_d^2 : 0 \leq \alpha \leq 1\}.$$

In all cases, the identified set is (weakly) smaller than the identified set under complete ignorance. More generally, it shrinks if confidence increases in the sense that each set Π_d shrinks; this happens, for example, if ϵ is reduced in the ϵ-contamination specification or in the variation neighborhood specification. It is easy to see also that an increase in confidence shrinks the set of empirical measures λ that can be rationalized by some Q. For example, in the absence of any confidence (complete ignorance), every λ with support in $\cup_d \{d(A) : A \in \mathcal{A}\}$ can be rationalized by some Q,[7] while in the ϵ-contamination specification λ can be rationalized only if it can be expressed as a mixture $(1-\epsilon)\widehat{\lambda}+\epsilon\widetilde{\lambda}$ where $\widehat{\lambda}$ is rationalizable under complete confidence ($\epsilon = 0$) and $\widetilde{\lambda}$ is rationalizable under zero confidence ($\epsilon = 1$).

2.3 A characterization

The main question to be addressed is "*which measures Q can rationalize λ given $\{\Pi_d\}$?*" We provide a comprehensive answer under the assumption that each Π_d is the core of a convex capacity, that is, for each d,[8]

$$\Pi_d = core(\nu_d), \text{ for some } \nu_d \text{ convex.} \tag{2.4}$$

Though limiting, (2.4) is of interest in light of the role of convex capacities, described in the introduction, both in decision theory (specifically in modeling beliefs under ambiguity) and in robust statistics. We note also that it is satisfied by all of the preceding specifications.

We employ the following notation. Fix a decision rule $d : \mathcal{A} \longrightarrow X$. Then, any distribution over menus $\pi_d \in \Delta(\mathcal{A})$, induces beliefs over alternatives $\pi_d \circ d^{-1} \in \Delta(X)$, where $\pi_d \circ d^{-1}$ denotes the probability measure given by

$$\left(\pi_d \circ d^{-1}\right)(a) = \pi_d\left(d^{-1}(a)\right) = \pi_d\left(\{A \in \mathcal{A} : d(A) = a\}\right). \tag{2.5}$$

Similarly, any capacity ν_d on the set $\mathcal{A}$ of menus induces the capacity $\nu_d \circ d^{-1}$ on the set X of alternatives, where

$$\left(\nu_d \circ d^{-1}\right)(K) = \nu_d(d^{-1}(K)) = \nu_d\left(\{A \in \mathcal{A} : d(A) \in K\}\right).$$

Under limited confidence about probabilities, if ν_d models the analyst's beliefs about menus conditional on decision rule d, then $\nu_d \circ d^{-1}$ models her induced beliefs about which alternatives are chosen by d.

Theorem 2.1. *Let $\{\Pi_d\}$ be such that, for each d, $\Pi_d = core(\nu_d)$ for some convex capacity ν_d on $\mathcal{A}$. Then $Q \in \Delta(\mathcal{D})$ rationalizes λ given $\{\Pi_d\}$ if and only if*

$$\lambda(K) \geq \sum_{d \in \mathcal{D}} Q(d) \cdot \left(\nu_d \circ d^{-1}\right)(K) \quad \textit{for all } K \subset X. \tag{2.6}$$

[7] A proof is given in the Supplementary Appendix.

[8] Since a convex capacity is uniquely determined by its core (see (A.3)), ν_d is necessarily unique. Appendix A provides all details about capacities needed in the sequel.

In particular, this equivalence applies to the five special cases of $\{\Pi_d\}$ *described above where the corresponding capacities* ν_d *are given by: For any collection of menus* $\mathcal{C} \subset \mathcal{A}$*,*

$$\begin{array}{lcl} ignorance & \nu_d(\mathcal{C}) = & \mathbf{1}_{\mathcal{A}}(\mathcal{C}) \\ contamination: & \nu_d(\mathcal{C}) = & (1-\epsilon)\,\widehat{\rho}_d(\mathcal{C}) + \epsilon \mathbf{1}_{\mathcal{A}}(\mathcal{C}) \\ variation\ nbhd & \nu_d(\mathcal{C}) = & \begin{array}{c} \max\{P_d(\mathcal{C}) - \epsilon, 0\} \ \mathit{if}\ \mathcal{C} \neq \mathcal{A} \\ and\ = 1\ \mathit{if}\ \mathcal{C} = \mathcal{A} \end{array} \\ interval\ beliefs & \nu_d(\mathcal{C}) = & \begin{array}{c} \max\{p_{*d}(\mathcal{C}), p_d^*(\mathcal{C}) - \beta_d\} \\ \beta_d = p_d^*(\mathcal{C}) - 1 \end{array} \\ 2\text{-dim}\ beliefs & \nu_d(\mathcal{C}) = & \min\{P_d^1(\mathcal{C}), P_d^2(\mathcal{C})\} \end{array}$$

The main message is that the sharp identified set of measures Q is the set of solutions Q to the finite set of linear inequalities (2.6), and constitutes a (convex) polytope. Accordingly, *the model is as tractable as the special cases of complete confidence and complete ignorance that have been studied in the literature.*

A proof follows.

Proof: We are given that, for each d, ν_d is a convex capacity on the set $\mathcal{A}$ of menus and $d : \mathcal{A} \longrightarrow X$. The following facts about convex capacities generalizes the well-known change-of-variable formula for probability measures:[9]

$$\nu_d \circ d^{-1} \text{ is a convex capacity on } X, \tag{2.7}$$

and its core satisfies

$$core\left(\nu_d \circ d^{-1}\right) = \left\{\pi_d \circ d^{-1} : \pi_d \in core\left(\nu_d\right)\right\}. \tag{2.8}$$

(See Appendix A for details.)

By (2.8) and the assumption that $core\left(\nu_d\right) = \Pi_d$, the rationalizability condition (2.1) amounts to the statement that

$$\lambda \in \sum_d Q(d)\ \ core\left(\nu_d \circ d^{-1}\right), \tag{2.9}$$

[9]If ν_d is a probability measure, say p_d, then $\nu_d \circ d^{-1} = p_d \circ d^{-1}$, both cores are singletons and (2.8) restates the preceding equation.

while condition (2.6) is the statement that

$$\lambda \in core\left(\sum_{d} Q(d)\left(\nu_d \circ d^{-1}\right)\right).$$

Because each capacity $\nu_d \circ d^{-1}$ is convex, the core of the mixture equals the mixture of the cores (see (A.4)), and hence the required equivalence with (2.6) is proven.

For the assertions regarding the special cases, one need only show that in each case the indicated capacity ν_d is convex and that it has core equal to the corresponding set Π_d. But these are well-known facts (Huber and Strassen 1973, Wasserman and Kadane 1990); for 2-dimensional beliefs, see also Topkis (1998, Lemma 2.6.4). ■

Remark: As explained in section 2.5, our theorem is "essentially" equivalent to Strassen's (1965) Theorem 4, when the latter is specialized so that all sets are finite. However, our proof using the mixture-linearity property of the core (A.4) is arguably much simpler. We view the greater simplicity as significant not as a mathematical contribution, but rather because it enhances transparency and accessibility of the theorem which, we believe, may help to expose and promote it as a useful tool for economists. Another value-added over Strassen is our demonstration of the theorem's usefulness.[10]

The following discussion provides additional perspective on the theorem and its value-added. Consider first the special case of complete ignorance. The associated capacities, written more fully, are given by

$$\nu_d(\mathcal{C}) = \begin{cases} 1 & \mathcal{C} = \mathcal{A} \\ 0 & \mathcal{C} \neq \mathcal{A} \end{cases}$$

The epistemic interpretation is certainty that (for every d) the menu is in $\mathcal{A}$, but there is complete ignorance within $\mathcal{A}$. The condition (2.6) characterizing rationalizability specializes to the set of inequalities

$$\lambda(K) \geq Q\left(\{d \in \mathcal{D} : \cup_{A \in \mathcal{A}} d(A) \subset K\}\right) \quad \text{for all } K \subset X. \tag{2.10}$$

Similar conditions have appeared previously in Barseghyan et al (2021, Theorem 3.1) and in Azrieli and Rehbeck (2025). As indicated in the introduction, the latter addresses different questions, and the former assumes a

[10] Doval and Eilat (2023) apply Strassen's Theorem 3, in addition to network flow arguments, in their proofs.

designated minimum size for menus. The ignorance special case admits alternative proofs. For one, the associated capacities ν_d are belief functions and hence the result for that case can be derived by using random set theory, which is the approach taken by Barsheghyan et al (2021). Alternatively, it follows immediately from the well-known structure of the core of a belief function (Dempster (1967) or Wasserman (1990, Theorem 2.1)). Moreover, these alternatives apply also to the ϵ-contamination specification since its ν_ds are belief functions. However, they do not apply when the ν_ds are convex but not belief functions, such as in the other three special cases or at the level of generality in the theorem.

To illustrate further the greater richness provided by admitting capacities that are convex rather than only belief functions, consider additional specifications that extend those given in the theorem. As defined above the ϵ-contamination specification models an analyst who is concerned that the focal measure may be contaminated by *any* probability measure. In some circumstances, however, only a subset of contaminations may be relevant (in statistics see Berger and Berliner (1986) and Moreno and Cano (1991), and in decision theory see Kopylov (2016)). For example, they might be restricted to lie in a variation neighborhood (2.3), thus leading to the following generalization of (2.2):

$$\Pi_d = (1-\epsilon)\,\widehat{\pi}_d + \epsilon \Pi'_d$$

where $\Pi'_d \subset \Delta(\mathcal{A})$ is defined as in (2.3), for some radius $\epsilon' \neq \epsilon$ about π'_d. Then Π_d equals the core of the convex capacity $(1-\epsilon)\,\widehat{\pi}_d + \epsilon\nu'_d$, where ν'_d is the convex capacity whose core is Π'_d, and thus is accommodated by the theorem. Similarly for the further generalization whereby

$$\Pi_d = (1-\epsilon)\,\widehat{\Pi}_d + \epsilon \Pi'_d,$$

where $\widehat{\Pi}_d$ is the variation neighborhood about $\widehat{\pi}_d$ with radius $\widehat{\epsilon}$ (with associated convex capacity $\widehat{\nu}_d$). The interpretation is that there are two focal measures $\widehat{\pi}_d$ and π'_d, but each is known (or believed to be valid) only up to small perturbations of size $\widehat{\epsilon}$ and ϵ' respectively, and where the two hypotheses have prior subjective probabilities $(1-\epsilon)$ and ϵ. Note that the theorem applies to this specification because, by the mixture linearity property (A.4),

$$\begin{aligned}
\Pi_d &= (1-\epsilon)\,\widehat{\Pi}_d + \epsilon \Pi'_d \\
&= (1-\epsilon)\,core\,(\widehat{\nu}_d) + \epsilon core\,(\nu'_d) \\
&= core\,((1-\epsilon)\,\widehat{\nu}_d + \epsilon\nu'_d) \equiv core\,(\nu_d)\,.
\end{aligned}$$

However, none of ν_d, $\widehat{\nu}_d$ and ν'_d is a belief function and therefore this specification is not covered by the random set theory approach in Barseghyan et al (2021).

Another illustrative special case is where each $\Pi_d = \Pi$ is given by the variation neighborhood (2.3) where the focal measure $P_d = P$ and the radii $\epsilon_d = \epsilon$ are common, and one thinks of ϵ as being small. This models an analyst (analyst 1) who believes that every decision rule faces "approximately" the same distribution over menus, but who allows for the possibility that small differences from the common focal measure may vary arbitrarily with d. This contrasts with analyst 2 who is certain that every d faces the identical distribution while possibly being uncertain (ambiguous) about the specific common distribution. Thus analyst 2 is perfectly certain that menu distributions are identical, while analyst 1 desires robustness against small differences. Each scenario is applicable to some settings. Models along the lines of analyst 2 have been adopted in several applied works where one can interpret the different menus as arising from feasibility rather than consideration (Tenn and Yun 2008, Tenn 2009, Conlon and Mortimer 2013, Lu 2022), and in the theoretical contribution by Azrieli and Rehbeck (2025, section 4). However, we suggest that only analyst 1 fits a setting where menus are subjective consideration sets in which case one would expect menus to depend on preference (or the decision rule), as in the applied papers by Goeree (2008), and Abaluck and Adams-Prassl (2021). For the reasons given at the end of section 2.1, analyst 2 is not accommodated in our model, specifically by our notion of rationalizability (2.1).[11] However, we can illustrate that the choice of which model (analyst 1 or analyst 2) to adopt can lead to different conclusions about identification. Suppose that the noted uncertainty of analyst 2 is also described by the set Π above, and say that $Q \in \Delta(\mathcal{D})$ rationalizes λ given Π if there exists $\pi \in \Pi$ such that

$$\lambda(a) = \sum_A \pi(A) \sum_d Q(d) \mathbf{1}_{d(A)=a}, \text{ for all } a \in X.$$

Then it is easy to see that the latter is equivalent to the modification of (2.1) where one adds the restriction $\pi_d = \pi_{d'}$ for all d and d'. Hence adoption of analyst 2 makes rationalizability more difficult and consequently shrinks the sharp identified set.

Application of the theorem requires that when considering whether to adopt a specification of interest for the Π_ds one is able to check whether it satisfies (2.4). For the particular specifications addressed in the theorem, the literature has confirmed (2.4). More generally, an important observation is that, given $\{\Pi_d\}$, then, for each d, there is only one candidate for a suitable capacity ν_d, namely the *lower probability* corresponding to Π_d and defined by

$$\nu_d(\mathcal{C}) = \inf\{\pi(\mathcal{C}) : \pi \in \Pi_d\}, \text{ for all } \mathcal{C} \subset \mathcal{A}.$$

[11] A referee made this point.

In other words, (2.4) is *equivalent* to the assumption that the lower probability capacity is convex and has Π_d as its core. (This follows directly from (A.3).) Convexity of ν_d can be checked, in principle, by using its definition (A.1) or any of its equivalent characterizations (Grabisch (2016, Theorem 3.15), for example). Since $\Pi_d \subset core\,(\nu_d)$ follows from the definition of lower probability, equality amounts to the requirement that Π_d be sufficiently large that it includes $core\,(\nu_d)$.[12]

A final related question is what can be done if one drops the assumption (2.4). It is straightforward to show that the counterpart of (2.6) given below is *necessary* for rationalizability given $\{\Pi_d\}$: If $Q \in \Delta(\mathcal{D})$ rationalizes λ, then, for the lower probability capacity ν_d and given $\pi_d \in \Pi_d$, (using $\Pi_d \subset core\,(\nu_d)$ and (A.5)),[13]

$$\lambda = \sum_d Q(d)\left(\pi_d \circ d^{-1}\right) \in \sum_d Q(d)\, core\left(\nu_d \circ d^{-1}\right)$$

$$\subset core\left(\sum_d Q(d)\left(\nu_d \circ d^{-1}\right)\right) \implies$$

$$\lambda(K) \geq \sum_d Q(d)\left(\nu_d \circ d^{-1}\right)(K) \quad \text{for all } K \subset X.$$

2.4 Another application: satisficing

Here we describe an application of our model where decision rules are not based on preference maximization. (The next subsection outlines an application that is not directly connected to choice.)

There is a population of satisficing decision makers each of whom chooses an alternative from the set X. They differ in two respects. First, aspiration thresholds differ; the set of distinct thresholds is $\mathcal{U}$. Second, individuals differ in the order in which they consider alternatives (this may be a subjective choice or exogenously imposed). Each sequential procedure follows a strict total order $>$ on X: the individual chooses the $>$-first alternative with a value at least as large as her threshold $u \in \mathcal{U}$, and if there are no such "satisfactory" alternatives then she chooses the $>$-last element in X. The empirical frequency of choices λ is observed, but both aspiration levels and orders $>$ are unobserved. Theorem 2.1, suitably reinterpreted, can be used to partially identify the distribution of aspiration levels while respecting limited

[12] Alternatively, given convexity, one can compute the cores by using the greedy algorithm Ichiishi (1981), or the algorithm in Chambers and Melkonyan (2005) that uses information about willingness to buy or sell and thus may help the analyst to calibrate parameters like ϵ.

[13] We use the trivial fact that $\pi_d \in core\,(\nu_d) \implies \pi_d \circ d^{-1} \in core\left(\nu_d \circ d^{-1}\right)$.

knowledge of the distribution of orders $>$. To see why, view an order as the counterpart of a menu and each threshold u, which maps each order into a chosen alternative, as a decision rule.

Similar applications can be made to other problems of choice with frames (Salant and Rubinstein 2008) where frames vary across individuals and are unobserved by the analyst.

2.5 A model variant

To conclude, we describe a slight modification of our model and then an application.

An initial motivation is relevant also to the choice setting addressed in our theorem. We have taken beliefs about menus, via $\{\Pi_d\}$, as a primitive and defined rationalizability relative to $\{\Pi_d\}$. However, because it is ultimately individuals' choices that matter for determining consistency with the data λ, an analyst might proceed by formulating beliefs directly on the space of alternatives. That calls for taking as a primitive a set $\mathcal{R}_d \subset \Delta(X)$ for each d and adopting the following alternative definition of rationalizability: Say that $Q \in \Delta(\mathcal{D})$ *rationalizes* λ *given* $\{\mathcal{R}_d\}$ *if, for all* d, *there exists* $\rho_d \in \mathcal{R}_d$ such that

$$\lambda(a) = \sum_{d \in \mathcal{D}} Q(d)\, \rho_d(a) \quad \text{for all } a \in X. \tag{2.11}$$

Q rationalizes λ given beliefs $\{\Pi_d\}$ over menus if and only if (2.11) is satisfied for the particular set of measures over alternatives $\mathcal{R}_d = \{\pi_d \circ d^{-1} : \pi_d \in \Pi_d\}$; hence our theorem characterizes (2.11) for those *particular sets* $\mathcal{R}_d$.

But we seek an analogue of Theorem 2.1 that applies to families $\{\mathcal{R}_d\}$ taken as primitives. In fact, with minor changes the proof of Theorem 2.1 yields the following characterization result: Let $\{\mathcal{R}_d\}$ be such that, for each d, $\mathcal{R}_d = core(\chi_d)$ for some convex capacity χ_d on X. Then $Q \in \Delta(\mathcal{D})$ rationalizes λ given $\{\mathcal{R}_d\}$ *if and only if*

$$\lambda(K) \geq \sum_{d} Q(d)\, \chi_d(K) \quad \text{for all } K \subset X. \tag{2.12}$$

(In fact, the proof is even simpler and consists *exclusively* of application of the mixture-linearity property of the cores of convex capacities. The mapping d from menus to alternatives, and its inverse d^{-1}, play no role in the current formulation. Because of this difference from Theorem 2.1, the above characterization result is *equivalent* to Strassen's (1965) Theorem 4 specialized so that all sets are finite.[14])

We conclude with an application that illustrates when the new notion of rationalizability may be particularly appealing.

Identifying the distribution of effort: Consider a population of workers with common observable characteristics (e.g. education and experience) who work independently. Each produces a homogeneous output in quantity represented by an element of X.[15] The empirical frequency distribution of outputs is given by $\lambda \in \Delta(X)$. Heterogeneity in output is attributed to differences in unobservable characteristics. The first unobservable is effort - there are finitely many effort levels $d \in \mathcal{D}$. The other unobservable is "everything else." The analyst may not be able to describe these other factors precisely, or even at all. However, we assume that she takes a stand on the set of their possible output consequences. Formally, for each effort d, the set $\mathcal{R}_d \subset \Delta(X)$ describes her view of the likelihoods of possible outcomes. As a special case, she views $O_d \subset X$ as the set of outputs possible given the effort level d and given what may ensue from "everything else." She "knows" the set O_d, but is ignorant about likelihoods *within* the set, that is, $\mathcal{R}_d = \Delta(O_d)$.[16]

With this reinterpretation, rationalizability of λ given $\{\mathcal{R}_d\}$ is well-defined, and (2.12) can be applied to yield the (computationally tractable) sharp identified set of measures Q over effort levels.

An alternative approach, that would rely on our earlier rationalizability notion (2.1), would be to introduce a parameter $\theta \in \Theta$ to represent "everything else," and, for each d, a fixed production function f such that the pair (d, θ) yields output $f(d, \theta)$. Beliefs about Θ would be captured by sets Π_d of distributions π_d over Θ in the counterpart of definition (2.1). (Roughly, Θ would play the role of the set of menus $\mathcal{A}$ above. With f presumed known, an effort d maps each θ into an output level; thus d acts like a decision rule.) However, such a formulation involving a production function f and probability distributions over Θ is arguably problematic in situations where the analyst cannot even conceive of what is included in "everything else."

[14]Our proof of (2.12) uses convexity of the χ_ds only to justify applying the mixture-linearity property of their cores, that is, the characterization provided by (2.6) is valid also if convexity is replaced by this mixture-linearity.Tijs and Branzei (2002) and Bloch and de Clippel (2010) give other assumptions, besides convexity, that imply mixture-linearity of cores, and our result applies immediately to them as well. The same cannot be said for Strassen's theorem. But it remains for future work to determine if any of the other assumptions provide alternatives to convexity that are interesting in our setting.

[15]To make clear the connection to the main choice model, we use the same symbols, though with different interpretations.

[16]The assumption that O_d can be specified even though "everything else" is poorly understood brings to mind Maskin and Tirole (1999) who argue that optimal contracts survive even with unforeseen *contingencies* when agents can forecast future *payoffs*.

The upshot of this section, and of the paper more broadly, may be summarized as follows. For the most part we have interpreted the paper as addressing heterogeneity in choice assuming heterogeneity in decision rules and the unobservability of menus. However, with suitable reinterpretations of the formal model, its scope is broader. Theorem 2.1 is relevant also to other contexts where one seeks the identification of a heterogeneous characteristic of prime interest that is robust to the distribution of other unobservables. In addition, where those secondary factors are not only unobservable, but also indescribable, or indefinable, we propose the modified notion of rationalizability (2.11) and its characterization (2.12) as an arguably superior approach. The distinction between the two formulations might be described by adapting Donald Rumsfeld's famous quote and referring to robustness to known unobservables (Theorem 2.1) versus robustness to unknown unobservables (via (2.11) and (2.12)).

A Appendix: Basic facts about capacities

Consider capacities on an arbitrary finite set Y; in the two special cases of interest Y equals the set of menus $\mathcal{A}$ or the set of alternatives X. Say that ν is a *capacity* on Y if $\nu : 2^Y \rightarrow [0,1]$, $\nu(\varnothing) = 0$, $\nu(Y) = 1$ and $\nu(K') \geq \nu(K)$ whenever K' is a superset of K. ν is *convex* if, for all subsets K' and K,

$$\nu(K' \cup K) + \nu(K' \cap K) \geq \nu(K') + \nu(K). \tag{A.1}$$

ν is a *belief function* if, for all n, and for all subsets $K_1, ..., K_n$,

$$\nu\left(\cup_{j=1}^{n} K_j\right) \geq \sum_{\varnothing \neq J \subset \{1,...,n\}} (-1)^{|J|+1} \nu\left(\cap_{j \in J} K_j\right). \tag{A.2}$$

If one restricts n to be 2, then one obtains the condition defining convexity. Hence every belief function is convex. (Convexity is sometimes referred to as monotonicity of order 2 while (A.2) is called infinite or total monotonicity.) A more transparent and equivalent definition of a belief function is that ν is induced by a random set.[17]

Let C be a subset of Y and ν a capacity on C (hence $\nu(C) = 1$). Then ν can be viewed also as a capacity on Y by identifying ν with the capacity ν' on Y defined by

$$\nu'(K) = \nu(K \cap C) \text{ for all } K \subset Y.$$

[17] Dempster (1967) and Nguyen (1978) are two early references describing the connection of random sets to belief functions. See also Nguyen (2006).

Further, ν' is convex if and only if ν is convex. We often identify ν and ν' and do not distinguish them notationally.

For any capacity ν on Y, its core is the set of all dominating probability measures, that is,

$$core\,(\nu) = \{p \in \Delta\,(Y) : p\,(K) \geq \nu\,(K) \text{ for all } K \subset Y\}\,.$$

If ν is convex, then its core is nonempty and ν can be recovered from its core as its lower bound or envelope:

$$\nu\,(K) = \min\{p(K) : p \in core\,(\nu)\}. \tag{A.3}$$

If $\nu = p$ is a probability measure, then it is convex and $core\,(\nu) = \{p\}$.

If ν and ν' are two convex capacities on Y, and if $0 \leq \alpha \leq 1$, then the mixture $\alpha\nu + (1-\alpha)\,\nu'$ is also a convex capacity and its core satisfies

$$core\,(\alpha\nu + (1-\alpha)\,\nu') = \alpha core\,(\nu) + (1-\alpha)\,core\,(\nu')\,. \tag{A.4}$$

(See Danilov and Koshevoy (2000, p. 9) or Grabisch (2016, p. 156).) This "mixture linearity" of the core is the key property that we exploit to prove our theorem. Elsewhere, (at the end of section 2.3), we also make use of the following weaker, and elementary, property that applies to any (not necessarily convex) capacities

$$core\,(\alpha\nu + (1-\alpha)\,\nu') \supset \alpha core\,(\nu) + (1-\alpha)\,core\,(\nu')\,. \tag{A.5}$$

Finally, consider the following assertions made in the proof of Theorem 2.1. We are given the convex capacity ν_d on $\mathcal{A}$, and $d : \mathcal{A} \longrightarrow X$. Define the capacity $\nu_d \circ d^{-1}$ on X. Then: $\nu_d \circ d^{-1}$ is convex, and

$$core\left(\nu_d \circ d^{-1}\right) = \left\{\pi \circ d^{-1} : \pi \in core\,(\nu_d)\right\} \equiv \mathcal{R}.$$

Proof: Convexity follows from verifying the defining inequalities (A.1). The inclusion $\mathcal{R} \subset core\,(\nu_d \circ d^{-1})$ is immediate. For the reverse inclusion, let $\{K_j\}$ be any chain of subsets of X. Then $\{d^{-1}(K_j)\}$ is a chain in $\mathcal{A}$. Since ν_d is convex, there exists $\pi^* \in core\,(\nu_d)$ such that $\pi^*\,(d^{-1}(K_j)) = \nu_d\,(d^{-1}\,(K_j))$ for all j (Choquet 1953). Thus $(\pi^* \circ d^{-1})\,(K_j) = \nu_d\,(K_j)$ for all j. Apply Grabisch (2016, Theorem 3.15) to conclude that $\mathcal{R} = core\,(\nu_d \circ d^{-1})$. ■

Acknowledgements: We are extremely indebted to Rohan Dutta who was instrumental throughout this project. We thank also Max Amarante and participants at the McGill applied micro breakfast for suggestions.
Funding: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

[1] Abaluck J. and Adams-Prassl A. What do consumers consider before they choose? identification from asymmetric demand responses. *Quart. J. Econ.* (2021) 136, 1611-1663.

[2] Azrieli Y. and Rehbeck J. Marginal stochastic choice. *Theoretical Econ.* (2025), online.

[3] Barseghyan L., Coughlin M., Molinari F., and Teitelbaum J.C. Heterogenous choice sets and preferences. *Econometrica* (2021) 89(5), 2015-2048.

[4] Berger J. and Berliner M. Robust Bayes and empirical analysis with epsilon-contaminated priors. *Ann. Statist.* (1986) 14, 461-486.

[5] Bloch F. and de Clippel G. Cores of combined games. *J. Econ. Theory* (2010) 145, 2424-2434.

[6] Camerer C. Individual decision-making, pp. 587-704 in *The Handbook of Experimental Economics 1*, J. Kagel and A. Roth eds. Princeton U. Press, 1995.

[7] Caplin A., Dean M., and Leahy J. Rational inattention, optimal consideration sets, and stochastic choice. *Rev. Econ. Stud.* (2019) 86, 1061-1094.

[8] Cattaneo M.D., Ma X., Masatlioglu Y., and Suleymanov E. A random attention model. *J. Pol. Econ.* (2020) 128, 2796-2836.

[9] Chambers R.G. and Melkonyan T.A. Eliciting the core of a supermodular capacity. *Econ. Theory* (2005) 26, 203-209.

[10] Choquet G. Theory of capacities. *Annales de l'institut Fourier* (1954) 5, 131-295.

[11] Conlon C.T. and Mortimer J.H. Demand estimation under incomplete product availability. *AEJ: Micro* (2013) 5, 1-30.

[12] Danilov V.I. and Koshevoy G.A. Cores of cooperative games, superdifferentials of functions, and the Minkowski difference of sets. *J. Math. Anal. Appl.* (2000) 247, 1-14.

[13] Dardanoni V., Manzini P., Mariotti M. and Tyson C.J. Inferring cognitive heterogeneity from aggregate choices. *Econometrica* (2020) 88(3), 1269-1296.

[14] Dempster A.P. A generalization of Bayesian inference. *J. Royal Statist. Soc.* (B) (1968) 30, 205-247.

[15] Dempster A.P. Upper and lower probabilities induced by multivalued mappings. *Ann. Math. Statist.* (1967) 38, 325-335.

[16] Doval L. and Eilat R. The core of Bayesian persuasion. arxiv 2023.

[17] Filiz-Ozbay E. and Masatlioglu Y. Progressive random choice. *J. Pol. Econ.* (2023) 131, 716-750.

[18] Galichon A. and Henry M. Set identification in models with multiple equilibria. *Rev. Econ. Stud.* (2011) 78, 1264-1298.

[19] Goeree M.S. Limited information and advertising in the U.S. personal computer industry. *Econometrica* (2008) 76, 1017-1074.

[20] Grabisch M. *Set Functions, Games and Capacities in Decision Making.* Springer, 2016.

[21] Huber P.J. Robust estimation of a location parameter. *Ann. Math. Statist.* (1964) 35(1), 73-101.

[22] Huber P.J. and Ronchetti E.M. *Robust Statistics.* 2nd ed. Wiley, 2009.

[23] Huber P.J. and Strassen V. Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* (1973) 1, 251-263.

[24] Ichiishi T. Super-modularity: applications to convex games and to the greedy algorithm for LP. *J. Econ. Theory* (1981) 25, 283-286.

[25] Kopylov I. Subjective probability, confidence and Bayesian updating. *Econ. Theory (2016) 62, 635-658.*

[26] Lu Z. Estimating multinomial choice models with unobserved choice sets. *J. Econometrics* (2022) 226, 368-398.

[27] Manski C. The structure of random utility models. *Theory and Decision* (1977) 8(3), 229-254.

[28] Manzini P. and Mariotti M. Stochastic choice and consideration sets. *Econometrica* (2014) 82, 1153-1176.

[29] Masatlioglu Y., Nakajima D., and Ozbay E. Revealed attention. *Amer. Econ. Rev.* (2012) 102, 2183-2205.

[30] Maskin E. and Tirole J. Unforeseen contingencies and incomplete contracts. *Rev. Econ. Stud.* (1999) 66, 83-114.

[31] McFadden, D.L. Conditional logit analysis of qualitative choice behavior, pp. 105-142 in *Frontiers of Econometrics,* ed. P. Zarembka. New York: Academic Press 1974.

[32] Moreno E. and Cano J. Robust Bayesian analysis with epsilon-contaminations partially known. *J. Royal Statist. Soc.* (B) (1991) 53, 143-155.

[33] Nguyen H.T. On random sets and belief functions. *J. Math. Anal. Appl.* (1978) 68, 531-542.

[34] Nguyen H.T. *An Introduction to Random Sets*. Chapman and Hall, 2006.

[35] Salant Y. and Rubinstein A. (A,f): Choice with frames. *Rev. Econ. Stud.* (2008) 75(4), 1287-1296.

[36] Schmeidler D. Subjective probability and expected utility without additivity. *Econometrica* (1989) 57(3), 571-587.

[37] Shafer G. *Mathematical Theory of Evidence*. Princeton U. Press, 1976.

[38] Strassen V. The existence of probability measures with given marginals. *Ann. Math. Statist.* (1965) 36, 81-95.

[39] Tenn S. Demand estimation under limited product availability. *Appl. Econ. Letters* (2009) 16, 465-468.

[40] Tenn S. and Yun J. Biases in demand analysis due to variation in retail distribution. *Int. J. Ind. Organ.* (2008) 26, 984-997.

[41] Tijs S. and Branzei R. Additive stable solutions on perfect cones of cooperative games. *Int. J. Game Theory* (2002) 31, 469-474.

[42] Topkis D.M. *Supermodularity and Complementarity*. Princeton U. Press, 1998.

[43] Wasserman L.A. Prior envelopes based on belief functions. *Ann. Statist.* (1990) 18(1), 454-464.

[44] Wasserman L.A. and Kadane J.B. Bayes' theorem for Choquet capacities. *Ann. Math. Statist.* (1990) 18, 1328-1339.